ANIL SETH

A fi TU

O nouă știință a conștienței

Scanned with OKEN Scanner

Mai vast e Creierul decât –
Tăria – confruntată
Ea nu prea greu încape-n Creier
Cu tine deodată

Emily Dickinson
(traducere de Veronica Porumbacu)

Scanned with OKEN Scanner

Cuprins

Prolog 1

I: NIVEL

1 Adevărata problemă 11
2 Măsurarea conștienței 33
3 Phi 60

II: CONȚINUT

4 Percepția în sens invers 77
5 Vrăjitorul din Odds 99
6 Contribuția privitorului 120

III: SINE

7 Delir 147
8 Așteaptă-te pe tine însuți 151
9 A fi un animal-mașină 175
10 Un pește în apă 200
11 Grade de libertate 213

IV: ALTELE

12 Dincolo de om 231
13 Minți-mașină 251

Epilog 273
Mulțumiri 281
Note 285
Referințe 323
Index 349

Scanned with OKEN Scanner

PROLOG

Acum cinci ani am încetat să exist, pentru a treia oară în viața mea. Treceam printr-o mică operație, iar creierul mi se îmbiba cu anestezic. Îmi amintesc senzațiile de întunecare, detașare, destrămare...

Anestezia generală diferă fundamental de somn. Așa și trebuie; dacă ai fi doar adormit, bisturiul chirurgului te-ar trezi instantaneu. Anestezia profundă seamănă, mai degrabă, cu unele situații catastrofale precum coma sau cu starea vegetativă, când conștiența lipsește cu desăvârșire. Sub anestezie profundă, activitatea electrică a creierului este anihilată aproape complet – iar asta nu se întâmplă niciodată în viața obișnuită, nici în stare de somn, nici în stare de veghe. Faptul că anesteziștii reușesc să inducă la nivel cerebral astfel de transformări – intrarea și ieșirea din starea de inconștiență – este un miracol al medicinei moderne. Este o transformare, un fel de magie: anestezia este arta de a transforma, practic, oamenii în obiecte.

Aceste obiecte, firește, redevin oameni. Și eu am revenit, amețit și dezorientat, dar clar *acolo*. Părea că timpul nu a trecut. Atunci când mă trezesc dintr-un somn profund, uneori nu-mi dau seama cât timp a trecut, dar mi-e clar că a trecut – percep continuitatea dintre conștiența mea anterioară somnului și conștiența de după. Sub anestezie generală, lucrurile stau altfel. Acest „sub" poate dura cinci minute, cinci ore, cinci ani – chiar și cincizeci. Iar „sub" nu este o exprimare tocmai corectă. Pur și simplu nu eram acolo – premoniție a uitării la care te condamnă moartea și, prin lipsa calităților, tulburător de reconfortantă.

Scanned with OKEN Scanner

Anestezia generală nu are efect doar asupra creierului sau a minții, ci și asupra conștienței*. Influențând delicatul echilibru electrochimic la nivelul circuitelor neuronale din cap, starea fundamentală care asigură existența în sine este abolită temporar. În acest proces rezidă unul dintre cele mai mari mistere ale științei, dar și ale filosofiei.

Cumva, în fiecare creier uman, activitatea combinată a miliarde de neuroni, fiecare dintre ei o mică mașinărie biologică, generează o experiență conștientă. Și nu una oarecare, ci o experiență conștientă personală, chiar acum, chiar aici. *Cum anume se întâmplă asta? Cum de trăim la persoana I singular?*

Păstrez o amintire din copilărie în care, privindu-mă în oglindă, am realizat, pentru prima dată, că acea experiență a mea dintr-un anume moment – cea de *a fi eu* – se va sfârși la un moment dat și că acel „Eu" va muri. Cred că aveam opt sau nouă ani și, ca orice amintire timpurie, nu poate fi considerată neapărat reală. Este posibil, însă, ca la acel moment să fi realizat, despre conștiența mea, că se poate sfârși, că trebuie să depindă, într-un fel, de materialul din care eram făcut – de materialitatea fizică a corpului și a creierului meu. Mi se pare că acest mister mă preocupă, într-un fel sau altul, încă de pe atunci.

În primii ani de studenție la Universitatea Cambridge, la începutul anilor '90, iubirea adolescentină pentru fizică și filosofie s-a transformat în fascinație pentru psihologie și neuroștiință, chiar dacă la acea vreme respectivele domenii păreau să evite, chiar să pună la colț ideea de conștiență. Cercetarea asociată cu doctoratul m-a dus și pe un lung dar neașteptat de prețios ocol prin inteligența artificială și robotică, iar după șase ani petrecuți la Institutul de

* Folosesc termenul „conștiență", nu „conștiință" deoarece în limba engleză conștiența/*consciousness* este, conform definiției de la pag 14, „orice formă de experiență subiectivă, oricare ar fi această experiență", iar conștiința/*conscience* se referă la valorile morale și la comportamentul ghidat de aceste valori (n. trad.).

Scanned with OKEN Scanner

Neuroștiință din San Diego, pe malurile Pacificului, am reușit, în final, să investighez direct fundamentele cerebrale ale conștienței. Acolo am lucrat cu laureatul Nobel Gerald Edelman – una dintre cele mai cunoscute figuri care a restabilit importanța investigării științifice a conștienței.

Timp de mai bine de zece ani am fost co-director al centrului de cercetare – Centrul Sackler pentru Știința Conștienței de la Universitatea din Sussex – cuibărit printre dealurile verzi din South Downs, lângă orașul Brighton. Centrul nostru de cercetare aduce împreună specialiști în neuroștiință, psihologi, psihiatri, specialiști în neuroimagistică, magicieni de realitate virtuală, matematicieni, filosofi, toți încercând să proiectăm cât mai multă lumină asupra fundamentelor cerebrale ale experienței conștiente.

Fie că ești om de știință sau nu, conștiența este un mister important. Pentru fiecare dintre noi, experiența conștientă înseamnă *tot ceea ce este*. Fără ea, nu există nimic: nici lume, nici sine, nici interior, nici exterior.

Să ne imaginăm că o versiune viitoare a mea, poate nu foarte îndepărtată, îți propune un târg – ocazia vieții. Pot să-ți înlocuiesc creierul cu un aparat care să-l reproducă în totalitate, astfel ca, din exterior, nimeni să nu vadă vreo diferență. Acest aparat vine cu nenumărate avantaje – nu se strică și, poate, te face să trăiești veșnic.

Dar există o condiție. Cât timp nici acel eu-din-viitor nu știe sigur cum anume generează creierul conștiența, nu pot garanta că vei avea vreo experiență conștientă dacă vei accepta oferta mea. Poate că vei avea, atunci când conștiența depinde doar de capacitatea funcțională, de puterea și de complexitatea circuitelor cerebrale, poate că nu vei avea, atunci când depinde de un anume material biologic – de

Scanned with OKEN Scanner

neuroni, de exemplu. Bineînțeles, cât timp aparatul-creier generează comportamentul cu care ești obișnuit, atunci când te întreb pe tine, noul-tu, dacă ești conștient, noul-tu va spune da. Dar dacă, în ciuda răspunsului afirmativ, viața – pentru *tine* – nu mai e la persoana I singular?

Bănuiesc, deci, că nu ai accepta propunerea mea. Fără conștiență, nu mai contează dacă trăiești încă cinci sau cinci sute de ani. În tot acest timp, *nimic nu ar putea să pară că ai fi tu.*

Lăsând filosofia deoparte, e clar că putem aprecia cu ușurință importanța practică a înțelegerii fundamentelor cerebrale ale conștienței. Anestezia generală se numără, evident, printre cele mai mari invenții din toate timpurile. Perturbări ale conștienței, mai puțin plăcute, chiar deranjante, pot apărea odată cu lezarea creierului și cu bolile mentale pentru tot mai mulți dintre noi, inclusiv pentru mine. Și, pentru fiecare dintre noi, experiențele conștiente se schimbă de-a lungul vieții, trecând de la confuzia înfloritoare și fremătătoare din timpul primilor ani ai copilăriei, prin aparenta claritate, probabil iluzorie și, cu siguranță, deloc universală a maturității, apoi glisarea finală – pentru unii, surprinzător de rapidă – către disoluția sinelui, pe măsură ce degenerarea neuronală se generalizează. Exiști în fiecare dintre etapele acestui proces, doar că ideea conform căreia am avea un singur sine (suflet?), același de-a lungul timpului, pare a fi total eronată. Într-adevăr, unul dintre aspectele cele mai evidente ale misterului conștienței este natura *sinelui*. Este conștiența posibilă fără conștiența de sine și, dacă da, contează în vreun fel?

Răspunsurile la întrebări atât de dificile au nenumărate implicații asupra felului în care gândim despre lume și viață. Când anume apare conștiența pe parcursul dezvoltării? La naștere sau înainte de naștere, în uter? Dar animalele care nu sunt oameni – și nu numai primatele și alte mamifere, ci orice ființă vie, cum ar fi caracatițele sau chiar organisme

Scanned with OKEN Scanner

mai simple, ca viermii sau bacteriile – au ele conștiență? Există o conștiență a existenței la *Escherichia coli* sau la un pește? Dar mașinile viitorului? Va trebui să ne preocupe nu numai puterea pe care noile forme de inteligență artificială o vor dobândi asupra noastră, ci și felul în care *noi* le abordăm pe *ele* din punct de vedere etic. Aceste probleme îmi amintesc de simpatia tulburătoare pe care am simțit-o față de Dave Bormann atunci când a distrus personalitatea lui HAL în *2001: O Odisee Spațială*, prin simpla extragere a procesoarelor de memorie, unul câte unul. Situația clonelor lui Ridley Scott din *Blade Runner* provoacă mai multă empatie, ceea ce demonstrează importanța naturii noastre de mașini vii pentru experiența de a fi sine conștient.

Această carte este despre neuroștiința conștienței: o încercare de a înțelege cum anume universul interior al unei experiențe subiective se înrudește cu procesele biologice și fizice care se desfășoară în creier și în corp și poate fi explicată prin acestea. Este proiectul care m-a captivat de-a lungul întregii mele cariere și cred că a ajuns acum într-un punct în care se conturează niște idei de răspunsuri.

Aceste contururi de idei se schimbă deja, chiar dramatic – cum anume gândim despre experiențele conștiente ale lumii din jur, precum și despre noi, în cadrul acestei lumi. Felul în care gândim despre conștiență se reflectă în fiecare aspect al vieții noastre. O știință a conștienței nu este nimic altceva decât o consemnare a ceea ce suntem, a ce înseamnă să fiu eu, sau să fii tu și de ce orice existență „este", în cele din urmă.

Povestea pe care urmează să v-o spun este o abordare personală, modelată de mulți ani de cercetare, contemplare și conversație. Din punctul meu de vedere, conștiența nu poate fi „rezolvată" în aceeași manieră în care a fost

Scanned with OKEN Scanner

decodificat genomul uman, sau consemnată schimbarea climatică globală. Misterele ei nu se vor dezvălui după o iluminare de tip *evrika* – așa cum ne place să credem că se relevă descoperirile științifice, numai că acesta este doar un mit.

Pentru mine, o știință a conștienței ar trebui să explice cum anume diversele proprietăți ale conștienței depind de, dar au și legătură cu operațiunile din „echipamentul umed (wetware)" situat în cutia noastră craniană. Scopul științei conștienței nu ar trebui să fie – cel puțin, nu cel fundamental – de a explica de ce conștiența face parte, în primul rând, din univers. Nici de a înțelege cum anume funcționează creierul în complexitatea sa, strecurând misterul conștienței sub preș. Ceea ce sper să reușesc să vă arăt este că, studiind proprietățile conștienței în termeni de mecanisme ale creierului și corpului, *de ce* și *cum*-urile metafizice profunde ale conștienței vor deveni, puțin câte puțin, mai accesibile.

Am folosit termenul de „echipament umed (wetware)" pentru a sublinia faptul că creierul nu este un computer făcut din carne, ci o mașinărie care funcționează pe bază de chimie și de rețele electrice. Orice creier este parte a unui corp viu, încastrat în, dar și interacționând, în același timp, cu mediul înconjurător – mediu care, în multe cazuri, conține alți creieri întrupați. Pentru a explica proprietățile conștienței în termeni de mecanisme biofizice este nevoie să înțelegem creierul – și mințile conștiente – drept sistem *întrupat* și *încorporat*.

În cele din urmă, aș vrea să rămâneți cu o nouă concepție despre sine – probabil cel mai semnificativ aspect al conștienței, pentru fiecare dintre noi.

Suntem obișnuiți să credem, încă din secolul al XVII-lea, de pe vremea lui Descartes, că celelalte animale nu ar avea conștiență de sine, deoarece nu au minți raționale care să le ghideze comportamentul. Ar fi „animale-mașină":

Scanned with OKEN Scanner

automate din carne, lipsite de capacitatea de a reflecta asupra propriei existențe.

Nu sunt de acord. În opinia mea, conștiența ține, mai degrabă, de supraviețuire decât de inteligență. Suntem sine conștient *tocmai pentru că* suntem animale-mașină. Voi construi acest caz, conform căruia experiențele de *a fi tu* sau de *a fi eu* provin din felul în care creierul prezice și controlează starea internă a corpului. Esența identității de sine nu rezidă nici în mintea rațională, nici în sufletul imaterial. Este un *proces* întrupat profund, un proces ce explică simpla senzație de a fi viu și care reprezintă baza experienței personale a sinelui pentru orice experiență conștientă. *A fi tu* se referă literar la corpul tău.

Această carte este împărțită în patru părți. În prima, explic abordarea mea referitoare la studiul științific al conștienței. Tot aici abordez și „nivelurile" de conștiență – cât de conștient poate fi cineva sau ceva – precum și progresele tentativelor de a „măsura" conștiența. Partea a doua se referă la „conținutul" conștient – de ce anume ești conștient, atunci când ești conștient. Partea a treia se concentrează spre interior, către sine și către toate diferitele experiențe pe care identitatea conștientă le comportă. Partea a patra și finală – „altele" – explorează ce anume afirmă această nouă înțelegere a conștienței despre celelalte animale și despre posibilitatea existenței mașinilor simțitoare. Către finalul cărții, veți înțelege că experiențele noastre conștiente despre lume și sine sunt forme de predicție făcute de către creier – „halucinații controlate" – care apar cu, prin și *datorită* corpului nostru viu.

În ciuda reputației pătate pe care o are printre specialiștii în neuroștiință, Sigmund Freud a avut dreptate referitor la multe subiecte. Privind înapoi în istoria științei, el a identificat

Scanned with OKEN Scanner

trei „contraargumente" împotriva importanței pe care și-o autoatribuie specia umană, fiecare reprezentând un pas înainte semnificativ și care a rezistat în timp. Primul este reprezentat de teoria heliocentrică, prin care Copernic a arătat că Terra gravitează în jurul Soarelui și nu invers. Astfel, ni s-a dezvăluit faptul că nu ne aflăm în centrul universului; suntem un fir de praf plutind pe undeva, un punct albastru-pal suspendat în abis. Urmează Darwin, care a arătat că avem un strămoș comun cu toate celelalte viețuitoare – revelare care, în mod uimitor, nu este acceptată nici acum de toată lumea. Total lipsit de modestie, Freud ne servește drept al treilea contraargument împotriva excepționalismului uman propria teorie despre mintea inconștientă, teorie care a subminat ideea conform căreia viețile noastre mentale s-ar afla sub control conștient, rațional. Chiar dacă nu a nimerit-o în ceea ce privește detaliile, Freud avea absolută dreptate atunci când a afirmat că o explicație naturală a minții și a conștienței va reprezenta o detronare, probabil finală, a speciei umane.

Aceste schimbări radicale ale manierei în care ne privim pe noi înșine sunt bine-venite. Cu fiecare pas pe care-l facem, aprofundând, descoperim noi motive de încântare și ne dezvoltăm capacitatea de a ne vedea pe noi înșine mai puțin *separați de*, cât *parte a* naturii.

Experiențele noastre conștiente sunt parte a naturii, la fel precum corpurile noastre, precum lumea noastră. Iar atunci când viața se va sfârși, conștiența se va sfârși și ea. Atunci când mă gândesc la asta, mă întorc în timp la experiența mea – *non*-experiența – anestezia. La impresia de uitare pe care mi-a creat-o, poate consolatoare – dar uitare, fără îndoială. În meditația sa asupra morții, scriitorul Julian Barnes o descrie perfect. Atunci când conștienta se sfârșește, nu mai e nimic – chiar *nimic* – de care să te temi.

Scanned with OKEN Scanner

I

NIVEL

Scanned with OKEN Scanner

I
Adevărata problemă

Ce este conștiența?

Pentru o ființă conștientă, există ceva anume ce *pare să fie* acea ființă. Ceva anume pare să fie eu, ceva anume pare să fie tu, probabil ceva anume pare să fie o oaie sau un delfin. Toate aceste ființe trec prin experiențe subiective. *Pare că ceva anume* este eu. Pare, aproape cu certitudine, că nu există nimic care să fie o bacterie, un fir de iarbă sau un robot-jucărie. Pare că aceste lucruri nu trec prin experiențe subiective: nu au univers interior, nici conștiență de vreun fel.

Acest mod de a vedea lucrurile este apropiat de cel al filosofului Thomas Nagel care, în 1974, a publicat un articol, legendar acum, intitulat „Cum e să fii liliac?", în care a susținut că, în timp ce omul nu poate trăi experiențele unui liliac, la fel, nici liliacul nu știe ce înseamnă, de fapt, să fie liliac.* Am apreciat mereu abordarea lui Nagel, deoarece evidențiază *fenomenologia:* proprietățile subiective ale experienței conștiente, cum ar fi, în cazul unei experiențe vizuale, forma, structura și calitățile asociate respectivei experiențe, comparativ cu proprietățile subiective ale unei experiențe emoționale sau ale uneia olfactive. În filosofie, aceste proprietăți poartă numele de *qualia*: nuanța de roșu a roșului, junghiul geloziei, durerea ascuțită sau zvâcnetul ritmic al unei dureri de dinți.

Pentru ca un organism să fie conștient, trebuie să dețină un fel de fenomenologie a lui însuși. *Orice* fel de experiență – *orice* proprietate fenomenologică – contează la fel de mult

* Acest articol este unul dintre cele mai cunoscute din domeniul filosofiei minții. După Nagel, „un organism are stări mentale conștiente numai și numai dacă există ceva anume care pare *să fie* acel organism - ceva *pentru* acel organism". Nagel (1974), p. 2

Scanned with OKEN Scanner

ca oricare alta. Oriunde este trăită o experiență, acolo există și fenomenologie. O ființă care trăiește doar o secundă va fi conștientă doar atât timp cât există ceva care pare să fie acea ființă, chiar dacă tot ceea ce i se întâmplă este doar o senzație trecătoare de durere sau de plăcere.

Este util să facem diferența dintre proprietățile fenomenologice ale conștienței și proprietățile sale *funcționale* și *comportamentale*. Acestea se referă la rolurile pe care conștiența le poate juca în operațiunile executate la nivelul minții și creierului fiecăruia, la comportamentele de care e capabil un organism, datorită faptului că are experiențe conștiente. Deși funcțiile și comportamentele asociate cu conștiența sunt subiecte importante, ele nu reprezintă terenul ideal pe care să căutăm definiții. Conștiența este, mai întâi de toate, despre experiența subiectivă – despre fenomenologie.

Această afirmație poate părea evidentă acum, dar nu a fost mereu așa. În trecut, a fi conștient era totuna cu a fi capabil să vorbești, cu a fi inteligent sau cu a manifesta un comportament de un anume fel. Dar conștiența nu depinde de comportamentul orientat către exterior, așa cum se poate observa în timpul somnului cu vise sau în cazul celor care trăiesc cu paralizie corporală totală. Pentru a afirma că limbajul este necesar conștienței ar trebui să considerăm că bebelușii, adulții care nu mai pot vorbi, precum și majoritatea animalelor non-umane sunt lipsite de conștiență. Iar gândirea abstractă complexă este doar o mică parte – posibil o parte clar umană – a existenței conștiente.

Câteva teorii cunoscute ale științei conștienței continuă să pună accent mai degrabă pe funcție și comportament decât pe fenomenologie. Dintre acestea, se remarcă teoria „spațiului de lucru global" (global workspace theory), dezvoltată pe parcursul mai multor ani, printre alții, de către psihologul Bernard Baars și de specialistul în neuroștiințe Stanilas Dehaene. Conform acestei teorii, conținutul

Scanned with OKEN Scanner

mental (percepții, gânduri, emoții etc.) este conștientizat atunci când primește acces la un „spațiu de lucru" care – anatomic vorbind – este distribuit în zonele frontale și parietale ale scoarței cerebrale. (Scoarța cerebrală, sau cortexul, este stratul extern, extrem de pliat, al creierului, format din neuroni strâns împachetați).* Atunci când conținutul mental ajunge în acest spațiu de lucru, devenim conștienți de respectivul conținut, care poate fi utilizat apoi în ghidarea comportamentului într-o manieră mult mai flexibilă decât în cazul unei percepții inconștiente. De exemplu, sunt conștient de paharul cu apă așezat pe masă în fața mea. Pot să-l ridic și să-i beau conținutul, să-l arunc peste tastele computerului (tentant), să scriu un poem despre el sau să-l duc înapoi la bucătărie, acum că am realizat că stă aici de câteva zile. Percepția inconștientă nu mi-ar permite acest grad de flexibilitate comportamentală.

O altă teorie cunoscută, numită a „gândirii de ordin superior", propune ideea conform căreia conținutul mental ajunge să fie conștientizat atunci când un proces cognitiv de ordin „superior" se orientează cumva către respectivul conținut, permițând conștientizarea lui. În această teorie, conștiența este strâns legată de procese precum metacogniția – „cogniția despre cogniție" – care, din nou, pune accent mai degrabă pe proprietățile funcționale decât pe fenomenologie (mai puțin, totuși, decât teoria spațiului de lucru global). Ca și teoria spațiului de lucru global, teoria gândirii de ordin superior evidențiază rolul lobilor frontali ca zone cheie pentru conștiență.

Deși aceste teorii sunt interesante și captivante, nu le voi comenta suplimentar în această carte, deoarece ambele accentuează aspectele funcționale și comportamentale

* Fiecare emisferă cerebrală are patru lobi. Cei frontali se află în dreptul frunții, cei parietali sunt în spatele celor frontali, de ambele părți ale creierului, cei occipitali se află în spatele craniului, iar cei temporali, de ambele părți ale capului, în spatele urechilor. Unii consideră sistemul limbic - situat adânc în creier - al cincilea lob.

Scanned with OKEN Scanner

ale conștienței, în timp ce abordarea mea pornește de la fenomenologie – din însăși experiența – și doar de la acel nivel va comenta despre funcție și comportament.

Definiția conștienței ca „orice formă de experiență subiectivă, oricare ar fi această experiență" este foarte simplă și poate fi considerată banală – ceea ce e foarte bine. Atunci când un fenomen complex nu este înțeles pe deplin, definițiile formulate precis, dar prematur pot fi constrângătoare și chiar derutante. Istoria științei a arătat de nenumărate ori că definițiile utile evoluează în tandem cu înțelegerea științifică, servind mai degrabă ca suport pentru progresul științific, decât ca punct de plecare sau ca finalitate în sine. În genetică, de exemplu, definiția „genei" s-a schimbat considerabil pe măsură ce biologia moleculară s-a rafinat. Tot așa, pe măsură ce cunoașterea noastră despre conștiență se dezvoltă, definiția – sau definițiile – conștienței evoluează, la rândul ei/lor. Dacă, pentru moment, suntem de acord că, atunci când vorbim de conștiență, ne referim la fenomenologie, atunci putem trece la următoarea întrebare.

Cum anume are loc conștientizarea? Care este legătura dintre experiențele conștiente și mașinăria biofizică din creierul și corpul fiecăruia? Care este legătura dintre aceste experiențe și atomii, quarcii, superstringurile sau orice altceva de acest fel din care este făcut, în cele din urmă, universul?

Formularea clasică a acestei întrebări este cunoscută drept „problema grea" a conștienței. Această expresie a fost formulată de filosoful australian David Chalmers, la începutul anilor '90, și a determinat întreaga agendă a abordării conceptului de conștiență încă de pe atunci. Iată cum descrie chiar filosoful:

Scanned with OKEN Scanner

Fără îndoială că anumite organisme sunt subiectele experienței. Cum anume aceste sisteme sunt subiecte ale experienței – aceasta este întrebarea care provoacă nedumerire. De ce, oare, atunci când sistemele noastre cognitive se implică în procesarea informației vizuale și auditive, avem experiențe vizuale și auditive: calitatea tonului de albastru, senzația de Do central? Cum putem explica de ce există ceva ce pare a menține o imagine mentală sau provoacă o emoție? Se știe bine că experiența provine dintr-un substrat fizic, numai că nu putem explica de ce și cum se formează. Cum de generează procesarea fizică o viață interioară atât de bogată? Obiectiv, pare total nerezonabil că se întâmplă așa ceva, numai că iată că se întâmplă.

Chalmers pune în balanță problema grea a conștienței cu așa-numita problemă ușoară – sau probleme ușoare – referitoare la explicarea manierei în care sistemele fizice, precum creierul, pot genera un număr infinit de proprietăți funcționale și comportamentale. Aceste proprietăți funcționale includ procesarea semnalelor senzoriale, selectarea acțiunilor și controlul comportamentului, atenția, generarea limbajului etc. Problema ușoară se referă la tot ceea ce ființele vii, ca noi, pot face și care poate fi descris în termeni de funcție – transformarea unui input în output – sau de comportament.

Bineînțeles, așa-numitele probleme ușoare nu sunt ușoare deloc. Pentru a le rezolva, specialiștii în neuroștiință au nevoie de zeci, poate sute de ani. Chalmers subliniază faptul că problemele ușoare sunt, în principiu, ușor de rezolvat – ceea ce nu se poate afirma despre problema grea. Mai precis, ideea lui Chalmers este că nu există un obstacol conceptual în calea problemelor ușoare, ceea ce ar permite, în final, explicații în termeni de mecanisme fizice. Din contră, pentru problema grea, pare că nu există o

Scanned with OKEN Scanner

astfel de explicație. (Pentru clarificare, un „mecanism" este un sistem alcătuit din părți care interacționează în mod obișnuit, generând diverse efecte.) Chiar și după rezolvarea tuturor problemelor ușoare, una câte una, problema grea va rămâne la fel. „Chiar și când vom fi reușit să explicăm executarea tuturor funcțiilor în proximitatea experienței – discriminarea perceptuală, categorisirea, accesul intern, raportul verbal – va persista o aceeași întrebare, fără de răspuns: *De ce execuția oricărei funcții este însoțită de trăirea unei experiențe?*"

Rădăcinile problemei grele se întind până în vremurile Greciei Antice, poate în timpuri și mai îndepărtate, dar sunt formulate clar abia în secolul al XVII-lea, când René Descartes a împărțit universul în *res cogitans*, ceea ce ține de minte, și *res extensa*, ceea ce ține de materie. Această separare a inaugurat filosofia dualismului, care a complicat și încețoșat apoi orice abordare despre conștiență. Confuziile generate au devenit evidente prin proliferarea diferitelor scheme filosofice dedicate conștienței.

Trageți aer în piept, urmează „ism"-ele.

Abordarea filosofică pe care o prefer, care este și presupunerea implicită a multor specialiști în neuroștiință, o reprezintă *fizicalismul*. Conform acestuia, universul este alcătuit din lucruri fizice, iar stările de conștiență sunt fie identice cu, fie derivate din anumite aranjamente ale acestor lucruri fizice. Unii filosofi folosesc termenul de *materialism* în loc de fizicalism, dar în contextul nostru pot fi considerate sinonime.

La polul opus fizicalismului se află *idealismul*. Asociat adesea cu George Berkeley, cardinal din secolul al XVIII-lea, idealismul consideră că mintea sau conștiența sunt cele care reprezintă sursa supremă a realității, nu lucrurile materiale sau materia. Problema nu e cum anume materia generează mintea, ci cum anume materia este generată de minte.

Scanned with OKEN Scanner

Situați ciudat, oarecum la mijloc, *dualiștii* ca Descartes cred că mintea/conștiența și materia fizică sunt substanțe sau moduri de existență separate, ridicând problema spinoasă a manierei în care acestea interacționează. În prezent, puțini sunt filosofii sau oamenii de știință care aderă la această viziune. Majoritatea oamenilor, mai ales în vest, preferă să se lase amăgită în continuare de dualism. Este seducătoare impresia că experiențele conștiente *par* non-fizice și alimentează „dualismul naiv" în care „impresia" stă la baza unor convingeri despre cum stau de fapt, lucrurile. Așa cum vom vedea pe parcursul acestei cărți, felul în care ne par nu este, însă, totuna cu felul în care sunt lucrurile, de fapt.

O variantă destul de influentă a fizicalismului este *funcționalismul*. Ca și fizicalismul, este o presupunere comună și adesea neenunțată de către mulți specialiști în neuroștiință. Mulți dintre cei care consideră fizicalismul ca fiind subînțeles, consideră și funcționalismul tot implicit. Alegerea mea este de a fi, însă, agnostic și ușor suspicios.

Funcționalismul este ideea conform căreia conștiența nu depinde de materialul din care e făcut un sistem (alcătuirea fizică), ci doar de ceea ce face sistemul, de funcțiile pe care le execută, de cum anume transformă intrările (inputs) în ieșiri (outputs). Funcționalismul se bazează pe impresia că mintea și conștiența sunt forme de procesare a informației, forme care pot fi implementate în creier, dar pentru care creierul biologic nu este absolut necesar.

Observați cum s-a strecurat aici, neanunțat, termenul de „procesare a informației" (așa cum s-a întâmplat și în citatul din Chalmers, de acum câteva pagini). Acest termen este atât de prevalent în discuțiile despre minte, creier și conștiență, încât e foarte simplu să-l accepți. Ceea ce ar fi o greșeală, deoarece sugestia că creierul „procesează informație" ascunde câteva afirmații foarte categorice. În funcție de cine face presupunerea, aceste afirmații fie sunt

Scanned with OKEN Scanner

formulate de tipul „creierul e un fel de computer, mintea (și conștiența) fiind software-ul“ (sau „mindware“-ul), sau se referă chiar la ceea ce *este* informația în sine. Toate aceste afirmații sunt periculoase. Creierul e foarte diferit de computer – de cel cu care suntem obișnuiți, cel puțin. Iar „ce anume *este* informația“ reprezintă o întrebare la fel de năucitoare ca „ce anume este conștiența“, așa cum vom vedea mai departe în această carte. Aceste îngrijorări explică suspiciunea mea față de funcționalism.

Luând funcționalismul de bun, așa cum procedează mulți, am putea deduce despre conștiență că ar putea fi *simulată* pe un computer. Să ne amintim că, pentru funcționaliști, conștiența depinde doar de ceea ce face un sistem, nu și de materialul din care e făcut sistemul. Asta înseamnă că, dacă identifici corect relațiile funcționale – adică dacă te asiguri că un sistem are cartografierea corectă a „relațiilor input-output“ – atunci asta ar fi suficient pentru generarea conștienței. Cu alte cuvinte, pentru funcționaliști, *simularea* coincide cu *instanțierea* – ceea ce se traduce cu manifestarea în ființă, în realitate.

Cât de rezonabilă este această concluzie? În unele cazuri, simularea este, într-adevăr, instanțiere. Un computer care joacă Go, cum este câștigătorul universal AlphaGo Zero al companiei de inteligență artificială DeepMind, *chiar joacă Go*. În multe alte situații, însă, este altfel. De exemplu, previziunile meteo. Simulările computerizate ale vremii, oricât ar fi de detaliate, nu sunt nici ploioase, nici secetoase. Este conștiența mai degrabă ca Go, sau mai degrabă ca vremea? Nu căutați un răspuns – nu există, cel puțin până acum. Este suficient să apreciem existența unei întrebări valide. De aceea sunt agnostic în ceea ce privește funcționalismul.

Mai avem două „isme“ și gata.

Primul este *panpsihismul* – ideea conform căreia conștiența este o proprietate fundamentală a universului, alături de alte proprietăți fundamentale cum ar fi masa/

Scanned with OKEN Scanner

energia și sarcina electrică; că este prezentă într-o anumită măsură peste tot și în toate. Uneori oamenii se amuză de panpsihism, care ar susține că pietrele și lingurile sunt conștiente într-o anumită măsură, așa cum suntem și noi, dar acestea sunt doar glume. Există versiuni mult mai sofisticare ale ideii, cu câteva dintre ele ne vom întâlni în capitolele următoare, dar principala problemă a panpsihismului nu este nebunia aparentă a ideii – unele idei care par nebunești se dovedesc a fi, în final, foarte reale sau, măcar, utile – ci faptul că nu explică, de fapt, nimic și nici nu conduce la formularea unor ipoteze testabile. Este o rezolvare superficială a unui mister formulat de problema grea, combinată cu orientarea științei conștienței către o fundătură empirică.

În cele din urmă, avem *misterianismul*, asociat cu filosoful Colin McGinn, după care ar exista o explicație exclusiv fizică a conștienței – o soluție completă la problema grea formulată de Chalmers – dar pe care noi, oamenii, nu o putem găsi sau nici măcar recunoaște, dacă ne-ar fi prezentată de extratereștri, deoarece nu suntem și nici nu vom fi suficient de deștepți pentru asta. Există o explicație fizică a conștienței, dar ne este la fel de inaccesibilă precum criptomoneda, unei broaște. Este inaccesibilă cognitiv din cauza limitărilor mentale caracteristice speciei umane.

Ce se poate spune despre misterianism? Întotdeauna vor exista lucruri pe care nu le vom înțelege niciodată, din cauza limitelor creierului și minții noastre. Puțini sunt cei care înțeleg, de exemplu, cum zboară un Airbus A380 (și, totuși, am călătorit cu plăcere într-unul, la întoarcerea acasă din Dubai). Cu siguranță există lucruri care vor rămâne inaccesibile cognitiv pentru majoritatea, chiar dacă sunt inteligibile în principiu, cum ar fi detaliile complicate ale teoriei stringurilor din fizică. Cât timp creierul este un sistem fizic cu resurse limitate, cât timp unii creieri par incapabili să înțeleagă anumite lucruri, pare fără putință

Scanned with OKEN Scanner

de tăgadă că există astfel de lucruri, pe care niciun om nu le va putea înțelege vreodată. Totuși, este nejustificat de pesimist să includem implicit conștiența în acest domeniu cețos al ignoranței specifice de specie.

Unul dintre cele mai frumoase aspecte ale metodei științifice constă în faptul că e cumulativă și incrementală. În prezent, mulți dintre noi putem înțelege lucruri care ar fi părut total de nedeslușit, *chiar și în principiu*, strămoșilor noștri, poate chiar oamenilor de știință și filosofilor de acum câteva zeci de ani. În timp, mistere după mistere au fost elucidate prin aplicarea sistematică a rațiunii și experimentului. Dacă misterianismul ar fi o opțiune serioasă, atunci putem să abandonăm totul și să ne ducem acasă. Hai să nu o facem.

Aceste „isme" descriu diferite tipuri de a gândi despre relația dintre conștiență și univers, ca întreg. Atunci când le cântărim punctele forte și pe cele slabe este important să recunoaștem faptul că ceea ce contează cel mai mult nu este neapărat „corectitudinea" modelelor, demonstrabile ca adevărate, cât utilitatea fiecăruia dintre ele în progresul înțelegerii noastre privind conștiența. De aceea prefer varianta funcțional agnostică a fizicalismului. Pentru mine, acesta este modelul mental cel mai pragmatic și productiv în abordarea științifică a conștienței. De asemenea, este cel mai onest din perspectivă intelectuală, cel puțin în ceea ce mă privește.

În ciuda atractivității sale, fizicalismul nu este acceptat de către toți cercetătorii din domeniul conștienței. Una dintre provocările cele mai comune este așa-numitul experiment de gândire „zombi". Aici, zombi nu sunt semi-cadavrele mâncătoare de creier din filme, ci „zombi filosofici". Dar trebuie să scăpăm de ei în același mod,

Scanned with OKEN Scanner

altminteri posibilitatea formulării unei explicații naturale, fizicaliste a conștienței e moartă înainte de a se naște.

Un zombi filosofic este o creatură care nu poate fi decelată de una conștientă, dar lipsită, totuși, de conștiență. Un Anil Seth zombi ar arăta ca mine, ar acționa ca mine, ar merge și ar vorbi ca mine, dar nimic nu va *fi* ca mine, în lipsa unui univers interior, a unor experiențe trăite. Întreabă-l pe zombi Anil dacă este conștient, iar el va răspunde „Da, sunt conștient". Zombi Anil va fi scris și eseuri despre neuroștiința conștienței, inclusiv gânduri despre relevanța chestionabilă a unor zombi filosofici pe acest subiect. Numai că nimic din toate acestea nu implică o experiență conștientă, de fapt.

De aceea, se presupune că ideea de zombi oferă un argument împotriva explicațiilor fizicaliste ale conștienței. Dacă-ți poți imagina un zombi, atunci poți concepe o lume care nu poate fi decelată de a noastră, dar în care nu există conștiență. Dacă poți concepe o astfel de lume, atunci conștienta nu poate fi un fenomen fizic.

Și iată de ce nu merge. Argumentul zombi, la fel ca multe alte experimente de gândire care tind către fizicalism, este un argument imaginabil, iar argumentele imaginabile sunt slabe prin definiție. La fel ca în cazul multor altor argumente de acest tip, plauzibilitatea lor este invers proporțională cunoașterii pe care o deținem.

Îți poți imagina un A380 zburând invers? Sigur că da. Un avion imens, zburând în marșarier. Este, un astfel de scenariu, *posibil?* Ei bine, cu cât știm mai multe în domeniul aerodinamicii și a ingineriei aeronautice, cu atât devine mai puțin posibil. În acest caz, chiar și o cunoaștere minimă din domeniile menționate ne ajută să înțelegem că avioanele nu pot zbura în marșarier. Pur și simplu *nu se poate.**

* Elicopterul poate zbura în marșarier, dar nu e avion. Am fost ciudat de fericit să aflu că originea cuvântului „elicopter" nu este combinația de „heli" și „copter", așa cum am presupus mereu, ci „helico" (spirala) și „pter" (aripă). Acum are mult mai mult înțeles.

Scanned with OKEN Scanner

La fel e și cu zombi. Pe de o parte, este banal să-ți imaginezi un zombi filosofic. Mă gândesc la o versiune a mea hălăduind prin preajmă, fără a avea nicio experiență conștientă. Dar chiar pot concepe așa ceva *cu adevărat?* Ceea ce ar trebui să fac, *de fapt,* este să iau în calcul capacitățile și limitările unei rețele vaste alcătuite din miliarde de neuroni și din miliarde de miliarde de sinapse (conexiunile dintre neuroni), ca să nu menționez și celulele gliale și concentrațiile de neurotransmițători, toate împachetate într-un corp interacționând cu o lume ce include alți creieri, în alte corpuri. Pot face asta? Poate cineva să facă asta? Mă îndoiesc. Ca și cu A380, cu cât știm mai multe despre creier și relațiile lui cu experiențele conștiente și cu comportamentul, cu atât mai puțin imaginabil devine un zombi.*

Dacă ceva anume poate fi imaginabil sau nu, este adesea o observație psihologică despre persoana care imaginează, nu o perspectivă asupra naturii realității. Aici stă slăbiciunea zombi. Ni se cere să ne imaginăm inimaginabilul, iar prin acest act de înțelegere iluzorie tragem concluzii despre limitele explicației fizicaliste.

Suntem acum pregătiți pentru a aborda ceea ce numesc eu *adevărata problemă a conștienței.* Am formulat această manieră de a gândi despre știința conștienței pe parcursul mai multor ani, asimilând și construind pe viziunile multor altora. Abordarea adevăratei probleme este, cred eu, soluția prin care o știință a conștienței poate să fie formulată cu succes.

* Creierul uman adult conține circa 86 de miliarde de neuroni și de circa o mie de ori mai multe conexiuni. Dacă numărăm o conexiune pe secundă, avem nevoie de 3 milioane de ani pentru a termina numărătoarea. Mai mult, e tot mai evident că fiecare neuron este capabil să execute el însuși funcții extrem de complexe.

Scanned with OKEN Scanner

Conform adevăratei probleme, principalele scopuri ale științei conștienței sunt *explicarea, previziunea* și *controlul* proprietăților fenomenologice ale experienței conștiente. Aceasta înseamnă a explica de ce o anumită experiență conștientă este așa cum este – de ce are proprietățile fenomenologice pe care le are – în termeni de mecanisme și procese fizice care se desfășoară în creier și în corp. Aceste explicații ar trebui să ne permită să prezicem când anume vor avea loc anumite experiențe subiective specifice și să le putem controla intervenind în mecanismele care le stau la bază. Pe scurt, abordarea adevăratei probleme necesită să explicăm *de ce* un anumit tipar de activitate cerebrală – sau de alt proces fizic – se potrivește unui anumit tip de experiență conștientă, nu numai să constatăm că asta se întâmplă.

Adevărata problemă se deosebește de problema grea deoarece nu explică – cel puțin, nu în primă fază – de ce și cum este conștiența parte a universului. Nu caută o soluție magică prin care să obțină conștiența din mecanisme (sau altminteri). Se deosebește și de problema (problemele) ușoară deoarece se concentrează mai degrabă pe fenomenologie decât pe funcție sau pe comportament. Nu bagă sub preș aspectele subiective ale conștienței. Și, ca urmare a accentului pus pe mecanisme și procese, adevărata problemă se aliniază în mod natural cu o perspectivă fizicalistă universală asupra relației dintre materie și minte.

Pentru a lămuri aceste deosebiri, să ne întrebăm cum anume abordările diferite ar putea tinde să explice experiența subiectivă a senzației de „roșu".

Din perspectiva problemei ușoare, provocarea este să explici toate proprietățile mecanice, funcționale și comportamentale asociate cu experimentarea roșului: cum anume lumina de o anumită lungime de undă activează sistemul vizual, condițiile în care facem afirmații de tipul „acel obiect este roșu", comportamentul obișnuit la observarea culorii roșii la semafor, cum anume obiectele roșii induc uneori răspunsuri emoționale de un anume tip etc.

Scanned with OKEN Scanner

Conceptual vorbind, orice explicație referitoare la de ce și cum proprietățile funcționale, mecaniciste și comportamentale ar trebui să fie însoțite de orice fel de fenomenologie – în acest caz, fenomenologia „roșului" – rămâne în afara abordării problemei ușoare. Existența experienței subiective, opusă non-experienței, este domeniul problemei grele. Indiferent de câtă informație mecanicistă deținem, este perfect justificabil să ne întrebăm „în regulă, dar de ce este asociat acest mecanism cu o experiență conștientă?" Dacă iei foarte în serios problema grea, vei bănui mereu existența unei verigi lipsă între explicațiile mecaniciste și experiența subiectivă a lui „văd roșu".

Adevărata problemă acceptă faptul că experiențele conștiente există și se concentrează în primul rând asupra proprietăților lor fenomenologice. De exemplu, experiențele roșului sunt vizuale, se asociază de obicei, dar nu întotdeauna, cu obiecte, par să fie proprietăți ale suprafețelor, au diferite niveluri de saturație, definesc o categorie printre alte experiențe de percepere a culorilor, deși pot varia ușor în cadrul unei categorii și tot așa. Important: acestea sunt, toate, proprietăți *ale experienței în sine*, nu – cel puțin, nu în primul rând – ale proprietăților funcționale sau comportamentale asociate cu acea experiență. Pentru adevărata problemă, provocarea constă în a explica, prezice și controla aceste proprietăți fenomenologice, în termeni de lucruri care se întâmplă în creier și corp. Ne-ar plăcea să știm ce e cu tiparele specifice de activitate cerebrală – cum e activitatea complexă în buclă din cortexul vizual* – care explică (prezice și controlează) cum anume o experiență, cum ar fi cea a roșului, decurge în modul ei specific și nu în alt mod. De ce nu e ca cea a albastrului sau a durerii de dinți sau a geloziei.

Explicație, predicție și control. Acestea sunt criteriile prin care sunt evaluate cele mai multe proiecte științifice,

* Cortexul vizual se află în lobul occipital, în partea posterioară a creierului.

Scanned with OKEN Scanner

indiferent de cât de complicat ar putea să pară, inițial, fenomenul studiat. Fizicienii au făcut eforturi semnificative pentru a desluși secretele universului – în explicarea, predicția și controlul proprietăților sale – dar sunt, încă, perplecși când vine vorba să explice din ce e făcut universul sau de ce acesta există. Tot așa, știința conștienței poate aduce contribuții semnificative în elucidarea proprietăților și a naturii experiențelor conștiente, fără a fi necesar să explice cum și de ce acestea sunt și parte a universului în care trăim.

Nu trebuie nici ca explicațiile științifice să ne confirme impresiile. În fizică, de notorietate este cazul mecanicii cuantice, o știință de-a dreptul contraintuitivă, dar universal acceptată ca fiind cea mai potrivită pentru a pricepe natura realității fizice. S-ar putea, la fel, ca o știință matură a conștienței să ne permită să explicăm, prezicem și controlăm proprietățile fenomenologice, fără senzația interioară de „da, așa e, *bineînțeles* că așa trebuie să fie!"

Foarte important, adevărata problemă a conștienței nu constă în acceptarea înfrângerii în fața problemei grele. Adevărata problemă vine după problema grea – indirect, dar după. Ca să înțelegem de ce e așa, vă prezint „corelatele neuronale ale conștienței".

Încă sunt uimit de cât de desconsiderată era știința conștienței, chiar și acum 30 de ani. În 1989, cu un an înainte de a-mi începe studențiа la Universitatea Cambridge, fiziologul-șef Stuart Sutherland scria: „Conștiența este un fenomen fascinant, dar inefabil. E imposibil de precizat ce anume este, ce face sau de ce s-a dezvoltat. Nu s-a scris nimic valoros pe acest subiect". Acest rezumat extrem de critic a apărut tocmai în *Dicționarul internațional de psihologie* și ilustrează atitudinea despre conștiență pe

Scanned with OKEN Scanner

care am întâlnit-o adesea la începutul parcursului meu educațional academic.

În alte părți, departe de Cambridge, dar de care nu știam la acea vreme, situația era mai favorabilă. Francis Crick (unul dintre descoperitorii structurii moleculare a ADN, alături de Rosalind Franklin și de James Watson) și colegul lui, Christof Koch, ambii stabiliți în San Diego, California, formulau ceea ce avea să devină metoda dominantă în dezvoltarea științei conștienței – cercetarea corelatelor neuronale ale conștienței.

Definiția standard a corelatului neuronal al conștienței (NCC = neural corelate of consciousness) este următoarea: „mecanismele neuronale minime care, împreună, sunt suficiente pentru orice percept conștient, indiferent de care ar fi acela". Abordarea NCC propune ideea că există un anumit tipar neuronal responsabil de fiecare și de orice experiență – cum ar fi vederea roșului. De fiecare dată când această activitate există, se va trăi experiența roșului, iar atunci când nu există, nu se va trăi acea experiență.

Marele merit al abordării NCC este acela că oferă o rețetă practică pentru cercetare. Pentru a identifica un NCC, se creează o situație în care subiecții pot avea, uneori, o anumită experiență conștientă, alteori, nu, cu mențiunea că respectivele condiții sunt pe cât posibil aceleași. Astfel, se poate compara activitatea cerebrală între cele două situații cu ajutorul tehnicilor de imagistică precum rezonanța magnetică funcțională (fMRI) sau electroencefalografia (EEG).* Activitatea cerebrală specifică unei stări de „conștiență" dezvăluie NCC dedicată acelei experiențe anume.

* fMRI măsoară un semnal generat de un component metabolic (oxigenul din sânge) în raport cu activitatea neuronală - oferă date cu înaltă definiție spațială, dar doar indirect legate de activitatea neuronală. EEG măsoară semnalele electrice generate de activitatea unor populații mari de neuroni situate aproape de scoarța cerebrală. EEG înregistrează mult mai fidel activitatea cerebrală comparativ cu fMRI, dar cu definiție spațială mult mai mică.

Scanned with OKEN Scanner

Fenomenul „rivalității binoculare“ este un exemplu grăitor: arătăm fiecărui ochi câte o imagine diferită – să zicem, o față umană ochiului stâng și o casă ochiului drept. Percepția conștientă nu se va exprima prin crearea unei himere față-casă, ci va oscila între față și casă, aprofundându-le pe rând. Inițial, vezi o casă, apoi o figură, din nou o casă... șamd. Percepția conștientă se schimbă chiar dacă inputul senzorial persistă. Analizând ce se întâmplă în creier, devine posibil să delimităm activitatea cerebrală care duce la percepția conștientă de activitatea cerebrală care duce la orice altceva determinat de inputul senzorial. Activitatea cerebrală care însoțește percepția conștientă permite identificarea NCC pentru acea percepție.

Strategia NCC a fost extrem de productivă pe parcursul multor ani, oferind rezultate fascinante, dar limitările nu au întârziat să apară. Una dintre probleme este dificultatea, uneori imposibilitatea de a decela NCC „adevărată“ de mai mulți factori care pot contribui, cei mai importanți fiind acele evenimente neuronale care sunt fie pregătitoare, fie consecințe ale acelor NCC în sine. În cazul rivalității binoculare, activitatea cerebrală care se desfășoară odată cu percepția conștientă poate înregistra atât procese ascendente (preparatorii), cum ar fi „orientarea atenției“, cât și procese descendente, comportamentul verbal de „raportare“ – de exprimare verbală a faptului că vezi o casă sau o față. Deși legate de desfășurarea evenimentelor percepției conștiente, mecanismele neuronale răspunzătoare de atenție și de exprimarea verbală – sau alte preparative și consecințe descendente – nu trebuie confundate cu cele care sunt răspunzătoare de percepția conștientă în sine.

Corelațiile nu sunt și *explicații*, însă – aceasta este problema mai profundă. Știm cu toții că simpla corelare nu coincide și cu stabilirea cauzei, dar este adevărat și că unele corelări nu-și găsesc explicația. Chiar și cu un design

Scanned with OKEN Scanner

experimental tot mai ingenios, cu tehnici de imagistică tot mai sofisticate, corelarea în sine nu are suficiente explicații. Din această perspectivă, strategia NNC și adevărata problemă sunt ca două boabe într-o păstaie. Dacă ne limităm la colectarea de corelații între lucruri care se întâmplă în creier și lucruri care se întâmplă în viața de zi cu zi, atunci nu e de mirare că vom suspecta mereu un vid de explicație între fizic și fenomenal. Dacă, în schimb, trecem dincolo de stabilirea de corelații pentru a descoperi explicații care conectează proprietățile mecanismelor neuronale cu proprietățile experiențelor subiective – așa cum susține adevărata problemă – acest vid se va umple, în cele din urmă. Când vom fi capabili să prezicem (și să explicăm și să controlăm) de ce experiența roșului este în felul în care este – nu a albastrului, nu a geloziei – misterul manifestării roșului va fi mai mic sau, poate, va dispărea.

Abordarea adevăratei probleme are ambiția ca, pe măsură ce construim punți explicative tot mai solide între fizic și fenomenologic, ideea intuitivă a problemei grele, conform căreia conștiența nu poate fi înțeleasă niciodată în termeni fizici, să se destrame până la dispariția într-un nor metafizic. Atunci când se va întâmpla, vom stăpâni o știință a experienței conștiente.

Ce anume justifică această ambiție? Luați în calcul cum anume înțelegerea științifică a *vieții* s-a maturizat pe parcursul ultimelor două secole.

Nu cu mult timp în urmă, viața părea cel puțin la fel de misterioasă precum ne pare azi conștiența. Cercetători și filosofi din diverse timpuri se îndoiau de faptul că mecanismele fizice sau chimice ar putea explica vreodată proprietatea de *a fi viu*. Diferența dintre viu și neviu, dintre

Scanned with OKEN Scanner

animat și neanimat părea atât de fundamentală încât se considera implauzibil să se găsească o legătură de tip mecanicist între ele.

Filosofia *vitalismului* s-a dezvoltat plenar în secolul al XIX-lea. Susținută de biologi de frunte precum Johannes Muller și Louis Pasteur, a persistat destul de mult în secolul al XX-lea. Vitaliștii credeau că proprietatea de a fi viu poate fi explicată doar prin intermediul unui ingredient special: o scânteie de viață, un *elan vital.* Acum știm că nu e nevoie de un ingredient special. Vitalismul nu mai e de actualitate, dispărând din cercurile științifice. Deși multe lucruri despre viață sunt încă necunoscute – cum anume funcționează o celulă, de exemplu – ideea că pentru a fi viu e nevoie de un ingredient supranatural și-a pierdut credibilitatea. Greșeala fatală a vitalismului a fost că a interpretat o pană de imaginație ca fiind o perspectivă aspra necesității. Este aceeași greșeală care caracterizează argumentul zombi.

Știința vieții a învins miopia vitalismului mulțumită concentrării pe progresul practic – pe un accent pus pe „adevăratele probleme" referitoare la ce anume înseamnă să fii viu. În ciuda pesimismului vitalismului, biologii au mers mai departe, descriind proprietățile sistemelor vii, explicând (prezicând și controlând, de asemenea), ulterior, fiecare dintre proprietăți în termeni de mecanisme fizice și chimice. Reproducerea, metabolismul, creșterea, autoreparare, dezvoltarea, autoreglarea homeostatică – toate au putut fi explicate mecanicist, atât individual, cât și împreună. Pe măsură ce s-au elucidat detaliile – și se elucidează în continuare –, nu numai că misterul lui „ce e viața" se destramă, dar și conceptul în sine al vieții s-a ramificat, astfel încât „a fi viu" nu mai este considerată o proprietate singulară, de tipul totul-sau-nimic. Au apărut zone gri, cea mai faimoasă fiind cea a virusurilor, iar acum, cea a organismelor sintetice și chiar cea a aglomerărilor de

Scanned with OKEN Scanner

picături fine de ulei, fiecare dintre ele având măcar o parte dintre caracteristicile sistemelor vii. Viața se apropie tot mai mult de natural devenind, astfel, și mai fascinantă.

Această paralelă vine atât cu o sursă de optimism, cât și cu o strategie practică de abordare a adevăratei probleme a conștienței.

Optimismul se referă la faptul că cercetătorii contemporani ai conștienței ar putea fi într-o situație asemănătoare celei cu care s-au confruntat biologii care studiau natura vieții cu câteva generații în urmă. Ceea ce trece ca fiind misterios *acum*, ar putea să nu fie *întotdeauna* misterios. Pe măsură ce explicăm diversele proprietăți ale conștienței în termeni de mecanisme care le susțin, poate că misterul fundamental al „cum anume se produce conștiența" se va destrăma, la fel cum s-a destrămat cel al „ce e viața".

Bineînțeles, paralela dintre viață și conștientă nu e perfectă. Evident, proprietățile vieții sunt descriptibile obiectiv, în timp ce scopurile explicative ale științei conștienței sunt subiective – ele există doar la persoana întâi. Totuși, această barieră nu este de netrecut; în mare, înseamnă că datele relevante, pentru că sunt subiective, sunt și mai greu de colectat.

Strategia esențială provine din constatarea că, la fel ca viața, conștiența nu este doar un fenomen singular. Mutând reflectoarele de pe viața ca mister unic, înfricoșător, biologii au renunțat să mai viseze la o soluție de tip "evrica", tot de tip unic. Mai degrabă au preferat să desfacă „problema" vieții într-un număr de procese înrudite, dar identificabile. Aplicând aceeași strategie și pentru conștiență, în această carte mă voi concentra pe *nivel*, *conținut* și *sine* ca proprietăți fundamentale a ceea ce înseamnă *să fii tu*. Procedând astfel, vom putea obține o imagine mulțumitoare a experienței conștiente în general.

Scanned with OKEN Scanner

Nivelul de conștiență se referă la „cât de conștienți suntem" – pe o scară de la lipsa cu desăvârșire a unei experiențe conștiente de orice fel, ca în cazul stării de comă sau de moarte cerebrală, și trecând prin stări de conștiență care însoțesc viața de fiecare zi în perioadele de veghe.

Conținutul de conștiență se referă la de ce anume suntem conștienți – imagini, sunete, mirosuri, emoții, stări, gânduri, convingeri care ne construiesc universul interior. Conținutul de conștiență este alcătuit din varii *percepții* – interpretări cerebrale ale semnalelor senzoriale care alcătuiesc, împreună, experiența noastră conștientă. (Percepția, așa cum vom vedea, poate fi atât conștientă, cât și inconștientă.)

Avem, apoi, *sinele* conștient – acea experiență specifică de *a fi tu* și tema principală a acestei cărți. Experiența de „a fi un sine" este un subset de conținuturi conștiente, care cuprind experiențele de a avea un anume corp, o anume perspectivă personală, un set de amintiri anume, precum și experiențe de stări, emoții și de „liber arbitru". Lumea sinelui este, probabil, acel aspect al conștienței de care ne agățăm cel mai strâns, atât de strâns încât devine tentant să confunzi conștiența de sine (experiența de a fi tu însuți) cu conștiența în sine (prezența oricărui fel de experiență subiectivă, de fenomenologie de orice fel).

Făcând aceste distincții, nu afirm că aceste aspecte ale conștienței sunt complet independente. De fapt, nici nu sunt, iar a ne da seama cum anume se înrudesc reprezintă o altă provocare semnificativă pentru știința conștienței.

Cu toate acestea, împărțirea adevăratei probleme a conștienței în acești termeni generici are numeroase beneficii. Explicând pe bucăți, se pot propune posibile mecanisme care să servească explicației, predicției și controlului. La fel de important, îndepărtează ideea limitată conform căreia conștiența este „un ceva" – un mister singular și intimidant care se poate sustrage analizei

Scanned with OKEN Scanner

științifice cu totul. Vom vedea, în schimb, cum anume proprietățile diferite ale conștienței se adună împreună în diferite moduri, de la o specie la alta și chiar între indivizi diferiți. Există tot atâtea moduri de a fi conștient câte organisme conștiente diferite există.

În cele din urmă, problema grea în sine poate dispărea, astfel că vom fi capabili să înțelegem conștiența ca fiind o prelungire a naturii, fără a fi nevoiți să adoptăm vreun „ism" arbitrar care să stabilească modul în care fenomenologia și fizica sunt înrudite.

Aceasta este promisiunea adevăratei probleme. Pentru a vedea unde ne duce, citiți în continuare.

Scanned with OKEN Scanner

2

Măsurarea conștienței

Cât de conștient ești, chiar în acest moment? Ce anume face diferența dintre a fi deloc conștient și a fi o bucată de carne vie, sau de siliciu neviu, fără un univers interior? Noi teorii și tehnologii permit cercetătorilor, pentru prima dată, să măsoare *niveluri* de conștiență. Pentru a înțelege cât mai bine această cercetare, să-i analizăm originile.

În Parisul de secol XVII exista o pivniță adâncă, întunecată și rece situată sub Observator, pe malul stâng al Senei. Această pivniță a avut un rol surprinzător în istoria științei – un rol care demonstrează importanța *măsurătorii* în evoluția cunoașterii.

Filosofii și oamenii de știință de la acea vreme, deși nu se chemau oameni de știință, încă, se întreceau în a dezvolta termometre cât mai precise și, astfel, să înțeleagă din punct de vedere fizic natura căldurii. Populare teorii „calorifice", conform cărora căldura este o substanță care putea curge în și din obiecte, căzeau în disgrație. Corectarea acestor teorii depindea de experimente precise, în care „fierbințeala" sau „răcoreala" obiectelor putea fi evaluată sistematic. Astfel de experimente aveau nevoie atât de mijloace de măsurare a orice ar fi fost „căldura", pe o scală pe care puteau fi comparate diferite valori. Întrecerea avea ca scop dezvoltarea unui *termometru* precis și a unei scale de *temperatură*. Dar cum se putea stabili precizia unui termometru, dacă nu în raport cu o scală bine validată? Și cum se putea dezvolta o scară de temperatură dacă nu exista deja un termometru precis?

Primul pas în rezolvarea acestei enigme era găsirea unui punct fix: o referință care nu se schimbă și despre care se

Scanned with OKEN Scanner

putea presupune că are o temperatură constantă. Chiar și așa ceva era o provocare. Eventuale variante promițătoare precum punctul de fierbere a apei depind de factori precum presiunea aerului, care presiune variază cu altitudinea și cu vremea, și chiar și de influențe subtile precum rugozitatea vasului de sticlă. Din astfel de cauze, pentru un timp, acea pivniță pariziană – cu a ei temperatură joasă, aparent constantă – părea o alegere rezonabilă pentru un punct fix de temperatură. (Aceasta nu a fost singura sugestie neobișnuită. Probabil că cea mai ciudată a fost punctul de topire a untului, a unui Joachim Dalencé.)

În cele din urmă, au fost inventate termometrele cu mercur, care au dus la înlocuirea teoriilor calorifice de către noua știință a termodinamicii – o revoluție asociată cu figuri legendare precum Ludwig Boltzmann și Lord Kelvin. În termodinamică, temperatura este o proprietate la scară largă a mișcării moleculelor într-o substanță; la modul specific, energia cinetică moleculară medie. Cu cât mișcarea este mai rapidă, cu atât temperatura este mai mare. „Căldura" devine energia transferată între două sisteme aflate la temperaturi diferite. Important, termodinamica a făcut mai mult decât doar să stabilească energia cinetică medie *corelată* cu temperatura – a propus ca asta să fie temperatura, de fapt. Înarmați cu această nouă teorie, oamenii de știință puteau vorbi acum despre temperatura de la suprafața soarelui și chiar să identifice un „zero absolut" la care mișcarea moleculară să înceteze, în teorie. Primele scale bazate pe măsurarea anumitor substanțe (Celsius, Fahrenheit) au fost înlocuite de o scală bazată pe proprietăți fizice asociate (Kelvin, după numele lordului). Baza fizică a temperaturii și căldurii nu mai e un mister.

Am citit această poveste pentru prima dată în cartea *Inventarea temperaturii*, de Hasok Chang, istoric la Colegiul Universitar din Londra. Până atunci, nu am apreciat niciodată cu adevărat cât de importante sunt măsurătorile

Scanned with OKEN Scanner

pentru progresul științific. Istoria termometriei și impactul ei asupra înțelegerii căldurii oferă un exemplu grăitor despre modul în care capacitatea de a face măsurători cantitative detaliate, pe o scală definită de puncte fixe, are puterea de a transforma ceva misterios în ceva inteligibil.

O astfel de abordare ar putea funcționa, oare, și pentru conștiență?

Uneori, filosofii discută despre un „conștientometru" ipotetic, capabil să determine dacă ceva – o altă persoană, un animal sau, poate, o mașină – este conștient/ă sau nu. La o conferință din anii '90, în apogeul discuțiilor problemei grele, David Chalmers s-a folosit de un vechi uscător de păr pe post de „conștientometru", pe care l-a îndreptat spre propriul cap, subliniind cât de important ar fi un astfel de obiect, dacă ar exista. Am îndrepta conștientometrul către ceva și am citi răspunsul. Gata cu misterul dimensiunilor cercului magic al conștienței.

Conform poveștii temperaturii, însă, valoarea măsurătorii stă nu numai în furnizarea unui răspuns de tip da sau nu despre prezența sau absența unei proprietăți, ci și în faptul că face posibile experimentele cantitative care au potențialul de a transforma înțelegerea științifică.

În cazul în care conștiența se dovedește a fi ceva asemănător cu temperatura – adică, dacă există măcar un singur proces fizic pe care se bazează și este identic cu „a fi conștient" – rezultatul ar fi spectaculos. Nu numai că am fi capabili să determinăm „cât de conștient" este cineva, dar vom putea discuta cu claritate despre anumite „niveluri" sau „grade" de conștiență și despre tipuri de conștientă, la mare depărtare de exemplul nostru umanoid, parohial.

Chiar dacă povestea conștienței se va dovedi altfel, mai puțin ca cea a temperaturii și mai degrabă ca a vieții, așa cum cred că va fi, capacitatea de a face măsurători precise

Scanned with OKEN Scanner

va rămâne, fără îndoială, un pas esențial în formularea de punți explicative; pentru explicare, prezicere și control al naturii experiențelor subiective. Oricare ar fi scenariul, măsurătorile transformă calitativul în cantitativ, vagul în precis.

Măsurătorile au și motivații practice. Arta anesteziei – aplicată pe mai mult de patru milioane de oameni zilnic – implică menținerea pacientului într-o stare temporară de uitare, dar cu atenție să nu i se administreze o doză prea mare. Un conștientometru precis, de încredere, ar avea, evident, mare valoare în setarea acestui echilibru delicat, mai ales de când anestezia este însoțită frecvent de blocarea agenților neuromusculari care induc paralizie temporară, permițând chirurgilor să-și facă treaba, fără să fie deranjați de reflexele musculare ale pacientului. Și, așa cum vom vedea curând, este nevoie urgentă de noi metode pentru a decide dacă într-un creier sever deteriorat mai există conștiență – cazul pacienților cu diagnostice teribile de tip „stare vegetativă" sau „stare de minimă conștiență".

De fapt, se folosesc de mulți ani monitoare care înregistrează conștiența cerebrală. Cel mai cunoscut este monitorul de „indice bispectral". În timp ce detaliile despre astfel de monitoare rămân ascunse în descrierile patentelor, conceptul fundamental este de a asambla măsurători electroencefalografice (EEG) dintr-un anumit domeniu de frecvență, care trebuie să rămână la anumite valori. Aceste înregistrări îl ajută pe medic pe parcursul anesteziei. Este o idee bună, dar monitoarele de indice bispectral sunt, încă, destul de controversate, parțial deoarece au existat situații în care ceea ce citeau nu se potrivea cu semnele comportamentale ale conștienței, cum ar fi mișcarea ochilor sau amintirea manevrelor chirurgicale făcute de medic în timpul operației. O problemă și mai complicată, când vine vorba de știința conștienței, este aceea că indicele bispectral nu se bazează pe nicio teorie clar formulată.

Scanned with OKEN Scanner

Pe parcursul ultimilor ani, a început să apară o nouă generație de monitoare de măsurat conștiența – nu în sălile de operații, ci în laboratoare de neuroștiință. Spre deosebire de monitoarele anterioare, cele noi se bazează pe înțelegerea teoretică tot mai clară a bazelor cerebrale ale conștienței și și-au dovedit deja eficiența în practică.

Măsurarea nivelului de conștiență la om nu este totuna cu constatarea dacă cineva e treaz sau adormit. Nivelul de conștiență nu este totuna cu nivelul de activare fiziologică. Deși ambele sunt corelate adesea, conștiența (conștientizarea) și starea de veghe (trezia) sunt diferite în multe feluri, ceea ce e suficient pentru a arăta că ele nu depind de același substrat biologic. Atunci când visezi, e clar că ești adormit, dar treci prin tot felul de experiențe conștiente. La cealaltă extremă se află condiții catastrofale precum starea vegetativă (cunoscută și sub denumirea de „sindromul de trezie non-responsivă"), în care o persoană trece ciclic de la starea de veghe la cea de somn, dar nu manifestă niciun semn comportamental de prezență conștientă: luminile sunt aprinse, dar nu e nimeni acasă. Figura de pe pagina următoare ilustrează relația dintre conștiență și prezență conștientă în mai multe cazuri – normale sau patologice.

Pentru a urmări nivelul de conștiență, avem nevoie să știm ce anume, în creier, susține conștiența, spre deosebire de ce anume susține starea de veghe. Ar putea fi numărul de neuroni implicați? Nu pare să fie așa. Cerebelul („creierul mic" situat la ceafă, sub cel mare) are de patru ori mai mulți neuroni decât tot restul creierului la un loc, dar nu pare implicat în conștientizare. Există o condiție rară, numită agenezăcerebelară, în care cerebelul nu se dezvoltă, dar subiecții respectivi duc vieți aparent normale. Nu ne putem îndoi, deci, că sunt conștienți.

Scanned with OKEN Scanner

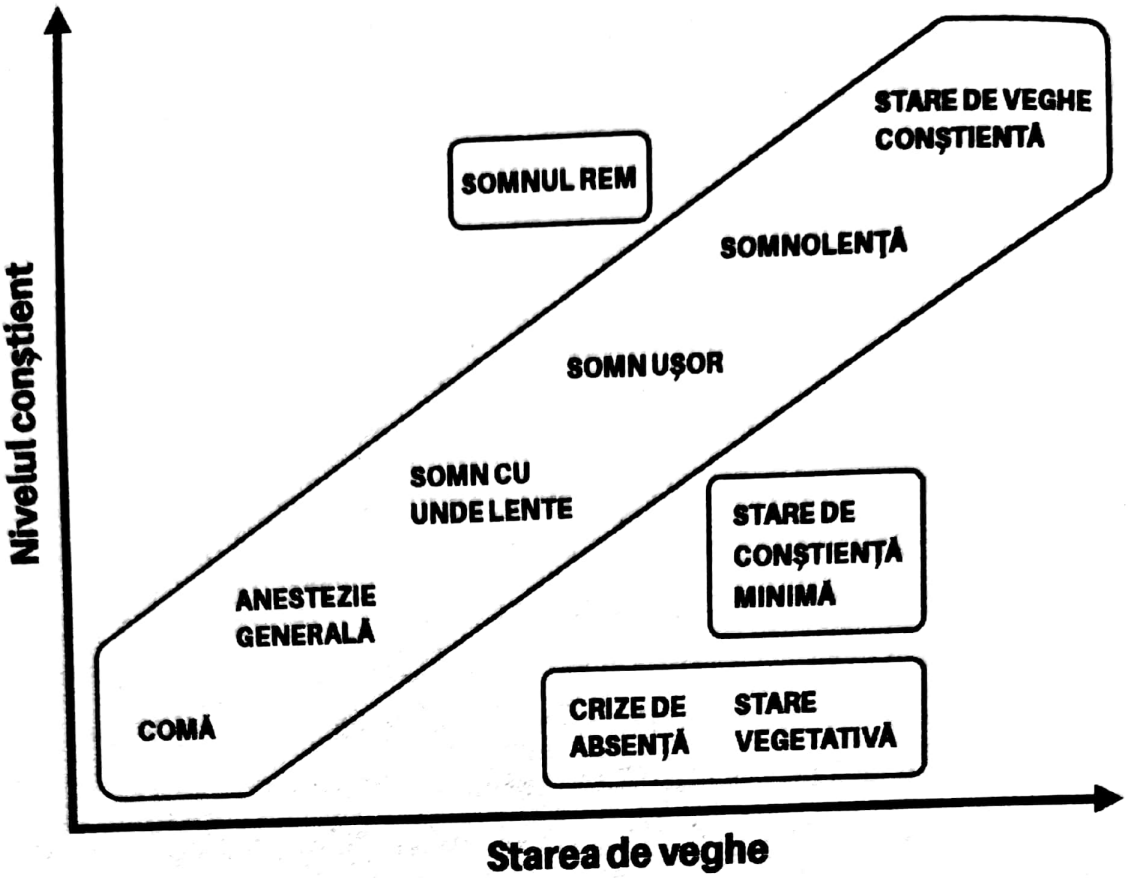

Fig. 1: Relația dintre nivelul de conștiență (prezența conștientă) și starea de veghe (activă).

Ce putem spune despre nivelul general de activitate neuronală? Este, creierul, la modul general, mai activ în stările de conștiență decât în cele de inconștiență? Poate că da, până la un punct – dar nu cu mult. Deși există diferențe în ce privește energia consumată la nivel cerebral pe diverse paliere de conștiență, aceste diferențe sunt, mai degrabă, mici, și nu poate fi vorba de o „oprire" a creierului atunci când conștiența se estompează.

În schimb, conștiența pare să depindă de maniera în care diferitele zone cerebrale comunică între ele, nu de creier ca întreg: tiparele de activitate care contează par să fie cele din sistemul talamo-cortical – combinația dintre cortexul cerebral și talamus (un set de structuri cerebrale ovale – „nuclei" – situate sub și conectate intim cu cortexul). Cele mai recente și interesante încercări de a măsura nivelurile de conștiență – și separarea lor de starea de veghe – se

Scanned with OKEN Scanner

bazează pe urmărirea și cuantificarea acestor interacțiuni. Cea mai ambițioasă versiune a acestei idei se soldează cu obținerea unui singur număr – care indică cât de conștientă este o persoană. Ceva asemănător cu un termometru.

Această abordare nouă a fost inițiată de Marcello Massimini, specialist italian în neuroștiință, la început împreună cu renumitul cercetător în domeniul conștienței, Giulio Tononi, de la Universitatea Wisconsin-Madison, apoi cu propriul său grup de la Universitatea din Milano. Ceea ce au făcut este simplu și elegant. Pentru a testa cum anume diferite părți ale cortexului discută una cu cealaltă, au stimulat activitatea dintr-o anumită zonă și au înregistrat, în timp și spațiu, cum anume se răspândește respectivul impuls la celelalte regiuni corticale. Pentru asta, au combinat EEG și stimularea magnetică transcranială (TMS = transcranial magnetic stimulation). Instalația de TMS constă într-un electromagnet foarte bine controlat, care permite generarea de pulsuri de energie scurte și puternice direct în creier, prin craniu, în timp ce EEG, în acest caz, este folosită pentru înregistrarea răspunsului creierului la această stimulare. E ca și cum ai lovi creierul cu un ciocan electric și ai asculta, apoi, ecoul.

Surprinzător, poate, subiecții nu prea sunt conștienți de această stimulare prin TMS, decât dacă provoacă reacții precum o mișcare (atunci când magnetul este plasat deasupra cortexului motor, care controlează mișcările) sau un flash vizual (o „fosfenă“ – care apare atunci când este stimulat cortexul vizual). Iar atunci când stimularea face ca mușchii feței sau ai scalpului să se contracte, evident că asta se simte. În general, însă, modificările activității cerebrale induse de TMS nu generează nicio schimbare notabilă în experiența conștientă. Poate că nu e chiar atât

Scanned with OKEN Scanner

de surprinzător – demonstrează că nu suntem conștienți de ceea ce fac propriii noștri neuroni. Și de ce am fi?

Chiar dacă nu simțim direct pulsurile TMS, Massimini și Tononi au arătat că ecourile lor electrice pot fi folosite pentru a distinge între diferitele niveluri de conștiență. În stări de inconștiență, cum ar fi somnul fără vise și anestezia generală, ecourile sunt foarte simple. Există un răspuns foarte puternic în acea parte a creierului care este stimulată, dar acest răspuns se atenuează rapid, precum undele provocate de aruncarea unei pietre într-un lac liniștit. În timpul stărilor de conștiență, însă, răspunsul este foarte diferit: un ecou tipic se răspândește pe toată suprafața corticală, dispărând și reapărând sub formă de tipare complexe. Complexitatea acestor tipare, în timp și spațiu, implică faptul că diferitele părți ale creierului – în particular, sistemul talamocortical – comunică unele cu altele într-o manieră mult mai sofisticată în timpul stărilor de conștiență decât în timpul celor de inconștiență.

În timp ce diferența dintre cele două variante este adesea ușor de văzut, doar privind datele, cu adevărat interesant la această cercetare este posibilitatea de a măsura complexitatea ecoului. I se poate asocia un număr, care indică *gradul* de complexitate. Abordarea se numește „zap și zip": se folosește TMS pentru stimularea – „zap" – cortexului și un algoritm de computer pentru a cuantifica – „zip" – răspunsul, ecoul electric, sub forma unui număr.

Algoritmul folosit pentru zip este același ca cel folosit pentru comprimarea (zip) a fotografiilor digitale în fișiere mai mici. Orice tipar, fie că e cel al unei poze din vacanța de vară, fie al unui ecou electric care se desfășoară în creier în timp și spațiu, poate fi reprezentat ca o succesiune de 0 și 1. Pentru orice secvență non-aleatorie se va genera o reprezentare comprimată, un șir de numere mult mai scurt, care poate fi folosit pentru regenerarea completă a originalului. Lungimea celei mai scurte reprezentări

Scanned with OKEN Scanner

comprimate posibile este cunoscută drept *complexitate algoritmică* a secvenței. Complexitatea algoritmică are valoare minimă, pentru o secvență complet predictibilă (cum ar fi o una constând doar din 1 sau din 0), valoare maximă, pentru o secvență complet aleatorie și o valoare medie, pentru secvențe care conțin zone predictibile. Algoritmul „zip" – care calculează ceea ce se numește „complexitatea Lempel-Ziv-Welch" sau LZW, pe scurt – este o modalitate obișnuită de a estima complexitatea algoritmică pentru orice secvență.

Din măsurarea ecourilor înregistrate în experimentele lor, Massimini și echipa sa au obținut ceea ce ei au numit *indicele de complexitate perturbațională* (perturbational complexity index = PCI). Acesta se folosește de complexitatea LZW pentru a oferi o măsură (indice) a/al complexității algoritmice a răspunsului creierului la o perturbare – pulsul TMS.

Ei și-au validat măsurătorile arătând că valorile PCI, în timpul stărilor de inconștiență, cum ar fi somnul fără vise și anestezia generală, erau clar coborâte comparativ cu valorile din timpul stării de veghe. Dincolo de faptul că această constatare e liniștitoare, adevărata putere a abordării PCI este aceea că definește o scală continuă, permițând delimitări tot mai fine. Într-un studiu de referință din 2013, echipa lui Massimini a măsurat valorile PCI la un mare număr de pacienți cu leziuni cerebrale, care aveau tulburări de conștiență. Au constatat că valorile PCI se corelau foarte bine cu nivelul de afectare – diagnosticat independent, de către neurologi. De exemplu, subiecții aflați în stare vegetativă, stare în care se presupune că nu există conștiență, deși ei sunt treji, au scoruri PCI mai scăzute decât cei care se află în stări de conștiență minimală, în care semnele comportamentale ale conștienței apar și dispar. Au putut chiar să separe valorile PCI care atestă conștiența de cele care sugerează lipsa ei.

Scanned with OKEN Scanner

În grupul meu de cercetare de la Universitatea din Sussex, am studiat metode similare de a evalua nivelurile de conștiență. În loc să folosim TMS pentru a injecta pulsuri de energie în cortex, am măsurat complexitatea algoritmică a activității cerebrale naturale în desfășurare – numită „spontană". O puteți vedea ca „zip" fără „zap". Într-o serie de studii conduse de colegul meu, Adam Barrett, și de fostul nostru doctorand Michaes Schartner, am descoperit despre complexitatea activității corticale spontane – așa cum e măsurată de EEG – că scade semnificativ, atât în etapele timpurii ale somnului, cât și în anestezie. Am descoperit, de asemenea, despre complexitate că este cam aceeași în timpul somnului REM (când predomină mișcările rapide ale ochilor = rapid eye movement), ceea ce are sens, deoarece în timpul somnului REM se întâmplă visele – iar visele sunt conștiente. Massimi și echipa sa au găsit același tipar al rezultatelor după măsurătorile PCI, ceea ce susține ipoteza conform căreia aceste măsurători urmăresc mai degrabă nivelurile de conștiență decât starea de veghe.

Capacitatea de a măsura nivelurile de conștiență independent de starea de veghe nu e importantă din punct de vedere științific, dar poate aduce modificări radicale în practica medicilor neurologi și pentru pacienții acestora. Studiul din 2013 al lui Massimini demonstrează deja că PCI poate distinge starea vegetativă de starea de minimă conștiență. Măsurători precum PCI sunt extrem de importante în acest context, pentru că nu se bazează pe comportamente vizibile în exterior. Simpla stare de veghe – excitarea fiziologică – este definită în termeni de comportament. În mediul clinic, neurologul trage concluzia că un individ este în stare de veghe atunci când acesta

Scanned with OKEN Scanner

În grupul meu de cercetare de la Universitatea din Sussex, am studiat metode similare de a evalua nivelurile de conștiență. În loc să folosim TMS pentru a injecta pulsuri de energie în cortex, am măsurat complexitatea algoritmică a activității cerebrale naturale în desfășurare – numită „spontană". O puteți vedea ca „zip" fără „zap". Într-o serie de studii conduse de colegul meu, Adam Barrett, și de fostul nostru doctorand Michaes Schartner, am descoperit despre complexitatea activității corticale spontane – așa cum e măsurată de EEG – că scade semnificativ, atât în etapele timpurii ale somnului, cât și în anestezie. Am descoperit, de asemenea, despre complexitate că este cam aceeași în timpul somnului REM (când predomină mișcările rapide ale ochilor = rapid eye movement), ceea ce are sens, deoarece în timpul somnului REM se întâmplă visele – iar visele sunt conștiente. Massimi și echipa sa au găsit același tipar al rezultatelor după măsurătorile PCI, ceea ce susține ipoteza conform căreia aceste măsurători urmăresc mai degrabă nivelurile de conștiență decât starea de veghe.

Capacitatea de a măsura nivelurile de conștiență independent de starea de veghe nu e importantă din punct de vedere științific, dar poate aduce modificări radicale în practica medicilor neurologi și pentru pacienții acestora. Studiul din 2013 al lui Massimini demonstrează deja că PCI poate distinge starea vegetativă de starea de minimă conștiență. Măsurători precum PCI sunt extrem de importante în acest context, pentru că nu se bazează pe comportamente vizibile în exterior. Simpla stare de veghe – excitarea fiziologică – este definită în termeni de comportament. În mediul clinic, neurologul trage concluzia că un individ este în stare de veghe atunci când acesta

Scanned with OKEN Scanner

răspunde la stimulare senzorială, precum un zgomot puternic sau o ciupitură a brațului. Conștiența, însă, este definită prin termeni de experiență subiectivă interioară, astfel că poate fi legată doar indirect de ceea ce se vede din afară.

Abordarea clinică standard pentru determinarea stării de conștiență a unui pacient cu creierul afectat se bazează, încă, pe comportament. În mod obișnuit, neurologul evaluează dacă un pacient răspunde nu numai la stimulare senzorială – dovada excitării fiziologice – dar este capabil și să interacționeze cu mediul înconjurător, reacționând la comenzi sau manifestând comportament voluntar. Dacă un pacient răspunde unei întrebări alcătuite din două părți și își știe atât numele, cât și data nașterii, atunci înseamnă că e pe deplin conștient. Problema cu această abordare este că unii pacienți pot avea viață undeva în interior, dar nu o pot exprima către exterior. Deducțiile care se bazează doar pe comportamente omit aceste cazuri, ajungându-se la diagnosticul de lipsă a conștienței când, de fapt, aceasta există.

Un exemplu extrem este „sindromul locked-in" (blocat în interior), în care conștiența este deplină într-un corp complet paralizat. Această situație rară poate fi consecința deteriorării trunchiului cerebral, o zonă situată la baza creierului (și la capătul superior al măduvei spinării) care, printre alte roluri, îl are și pe acela de a media controlul mușchilor corpului și de la nivelul feței. Printr-un capriciu anatomic, pacienții locked-in/blocați în interior pot face mișcări limitate ale ochilor, ceea ce permite manifestarea unui comportament – greu de neobservat, uneori – care ar permite diagnosticare și comunicare. Un fost editor al revistei *Elle* – Jean-Dominique Bauby, care a ajuns locked-in în 1995 din cauza unei hemoragii cerebrale – a scris o carte în acest fel, *The Diving Bell and the Butterfly (Clopotul de scufundare și fluturele)*. Pacienții total locked-in nu pot să folosească, însă, nici aceste canale de comunicare, diagnosticul fiind și mai greu de pus. Atunci

Scanned with OKEN Scanner

când ne bazăm doar pe comportament, putem confunda ușor sindromul locked-in cu absența totală și permanentă a conștienței. Scanarea creierului la cei ca Bauby arată că, la nivel general, activitatea cerebrală este aproape normală. Massimini arată, în studiul lui din 2013, că pacienții locked-in aveau valori ale PCI similare cu ale subiecților sănătoși – dovedind o conștiență intactă.

Cele mai spectaculoase cazuri se situează în zona gri dintre viață și moarte, în stări ca cea vegetativă sau cea de conștiență minimă. În aceste zone de graniță, semnele comportamentale ale conștienței pot lipsi cu desăvârșire sau pot fi foarte rare, iar deteriorarea creierului, atât de extinsă, încât scanarea lui să nu poată oferi date concludente. În astfel de situații, măsurători ca PCI pot aduce transformări radicale. Atunci când un pacient are un scor PCI care sugerează conștiența, chiar și atunci când orice altceva afirmă contrariul, acel pacient merită o șansă suplimentară.

Marcello Massimini mi-a povestit recent despre un caz în care măsurarea PCI chiar a făcut diferența. Un tânăr cu răni grave la cap a fost internat într-un spital din Milano. Nu a răspuns nici la cele mai simple întrebări și cerințe, fiind diagnosticat cu stare vegetativă. Dar valoarea PCI era mare, ca la o persoană sănătoasă și perfect conștientă – observație șocantă, mai ales că nu era locked-in. Echipa de clinicieni a reușit să găsească o rudă, un unchi care a venit în Italia din nordul Africii, unde locuia familia tânărului. Atunci când acest unchi a început să vorbească nepotului în arabă, răspunsul lui a fost imediat: zâmbea la glume, chiar a ridicat degetul mare (thumb-up) vizionând un film. Fusese conștient tot acest timp – doar nu știa să răspundă *în italiană*. E greu de înțeles de ce s-a întâmplat astfel. Massimini crede că a fost un caz ciudat de „ignorare culturală", ca și cum lumea italiană nu ar mai fi contat deloc pentru el. Indiferent de explicație, sfârșitul poveștii acestui tânăr

Scanned with OKEN Scanner

ar fi fost cu totul altul dacă nu i s-ar fi măsurat PCI prin intermediul ecourilor electrice indicatoare.

Diagnosticul conștienței reziduale la pacienții cu creierul lezat este un domeniu medical în plină schimbare. Odată cu PCI al lui Massimini, se dezvoltă mai multe alte metode clinice preluate din laborator. Preferata mea este cea bazată pe experimentul „tenis / casă" – faimos în cercurile neurologilor – condus de Adrian Owen și echipa sa în 2006. Experimentul a constat în plasarea într-un scanner RMNf a unei tinere de 23 de ani cu lipsă a răspunsului comportamental, după un accident în trafic, și căreia i s-au dat mai multe instrucțiuni verbale. Uneori i se spunea să-și imagineze că joacă tenis, alteori, că se plimbă prin camerele casei ei. Aparent, pare ceva mult prea specific, cât timp astfel de pacienți nu răspund la nimic, mai ales la instrucțiuni verbale complexe. Totuși, studiile pe subiecți sănătoși au arătat că regiunile cerebrale implicate în imaginarea mișcărilor fluide (cum ar fi jocul de tenis) sunt cu totul altele decât cele activate de imaginarea deplasării prin diverse spații.*

Remarcabil a fost că pacienta lui Owen a manifestat exact același tipar de răspuns cerebral, demonstrând că și ea urmărea activ instrucțiunile, angajându-se în sarcina mentală de a-și imagina anumite acțiuni. Este aproape imposibil de conceput ca cineva să poată face asta atunci când este inconștient, astfel că Owen a declarat diagnosticul comportamental de stare vegetativă ca fiind eronat, iar tânăra, de fapt, era conștientă. Astfel, Owen și echipa sa au recalibrat un scanner cerebral pentru a permite unui pacient să interacționeze cu mediul extern prin intermediul creierului, nu al corpului.

* Imaginarea (și execuția) mișcărilor fluide activează regiuni corticale cum ar fi zona motorie suplimentară, în timp ce imaginarea deplasării prin diverse spații activează alte regiuni, cum ar fi girul parahipocampic. Anatomic, aceste arii cerebrale sunt situate la distanță una de alta. Evident, în ambele cazuri de imaginare, se activează regiuni implicate în procesarea sunetelor și a limbajului.

Scanned with OKEN Scanner

Ulterior, au început și alte studii după metoda lui Owen – nu numai pentru diagnostic, ci și pentru comunicare. Într-un studiu din 2010 al lui Martin Monti, un pacient internat cu diagnostic de stare vegetativă a fost capabil să răspundă cu da/nu la întrebări, imaginându-și că joacă tenis pentru „da“ și că se plimbă prin casă pentru „nu“. O metodă laborioasă de comunicare, cu siguranță, dar o soluție extraordinară pentru cei care nu mai au nicio posibilitate să se facă înțeleși.

Câte persoane non-reactive dar conștiente or fi zăcând uitate prin saloane de neurologie sau prin cămine? Greu de știut. Metoda lui Owen – mai veche decât PCI – a fost investigată mai des, o analiză recentă sugerând că între 10 și 20% dintre pacienții aflați în stare vegetativă ar putea avea o formă de conștiență ascunsă – procente care s-ar traduce în multe mii de oameni în toată lumea. Și chiar și așa, ar putea fi o subestimare. Pentru a trece de testul Owen, pacienții trebuie să înțeleagă limba și să fie capabili să-și imagineze diverse, ceea ce e imposibil pentru unii, chiar dacă sunt conștienți. Acestea sunt cazurile în care metode precum PCI devin importante, deoarece vin cu speranța detectării unei conștiențe reziduale, fără a-i cere pacientului să facă nimic special pentru asta. Exact cum ar trebui să facă orice conștientometru adevărat.

Conceptul de „nivel de conștiență“, așa cum l-am folosit până acum, se referă la schimbări globale relative în starea de conștiență a unui individ – cum ar fi diferența dintre viața normală, în stare de veghe, și anestezia generală sau starea vegetativă. Totuși, sunt și alte moduri de a gândi ce anume înseamnă nivelul de conștiență. Este un copilaș mai puțin conștient decât un adult? Este o țestoasă mai puțin conștientă decât altele?

Scanned with OKEN Scanner

Există, cu siguranță, un pericol în a gândi în acești termeni. Astfel de întrebări ne farmecă, inducându-ne ideea că orice formă de conștiență care diferă de cea a unui adult sănătos e, oarecum, inferioară. Este modul de gândire simptomatic al excepționalismului speciei umane, care a afectat biologia, întunecând istoria gândirii peste tot. Conștiența are nenumărate proprietăți și este o greșeală să confunzi expresia unui anumit set de proprietăți, tipice omului adult sănătos, cu esența conștienței în toate formele sale, apoi să presupui că aceasta e în vârful unei scale unidimensionale. Experiențele conștiente apar, cu siguranță, de-a lungul timpului, fie în dezvoltarea unui anume animal (om sau nu), fie în cadrul vastei expansiuni a evoluției. Dar este cu totul altceva să descrii oricare dintre aceste procese ca desfășurându-se liniar, sau culminând cu idealul de a fi eu sau tu al omului adult. Aceasta este una dintre situațiile în care analogia dintre conștiență și temperatură – cu care am început acest capitol – poate fi limitată.

O întrebare din același registru: este conștiența ceva de tip „tot sau nimic" – lumina e fie aprinsă, fie stinsă – sau e „graduală", fără o separare clară între conștiență și lipsa ei? Această întrebare se referă atât la apariția conștienței în decursul evoluției și al dezvoltării, cât și la întoarcerea din uitarea anesteziei sau a somnului fără vise. Deși ispititoare, cred că această întrebare este eronată. Diferența dintre conștiența „tot sau nimic" și cea „graduală" nu trebuie să fie radicală. Fie că vorbim de evoluție, de dezvoltare, de viața de zi cu zi sau de salonul de neurologie, prefer să gândesc în termeni de tranziție netă de la absența totală a conștienței, la prezența a cât de puțină experiență conștientă, apoi experiență conștientă manifestată în diferite grade, poate în diferite dimensiuni, odată ce luminile interioare se întrevăd, măcar.

Fie un om adult tipic. Este nivelul lui de conștiență mai înalt (sau mai coborât) atunci când visează decât atunci

Scanned with OKEN Scanner

când e la birou, după un prânz copios, într-o stare detașată, de semi-stupoare? Nu există răspunsuri clare la astfel de întrebări. Visatul ar putea fi „mai conștient" din anumite puncte de vedere (de exemplu, realismul fenomenologiei perceptuale), dar „mai puțin conștient", din altele (de exemplu, gradul de examinare interioară reflectivă a ceea ce se întâmplă*).

Dacă luăm în serios ideea de niveluri multidimensionale ale conștienței, atunci dispare distincția foarte clară dintre *nivelul* de conștiență și *conținutul* de conștiență. Devine lipsit de sens să separi complet cât de conștient ești de de ce anume ești conștient. O măsurătoare a conștienței de tip „o mărime universală" – de genul la care ne putem aștepta dacă luăm analogia cu temperatura prea literal – ar putea să nu fie de ajuns niciodată.

Un exemplu despre cum interacționează nivelul de conștiență cu conținutul de conștiență provine dintr-un studiu realizat cu câțiva ani în urmă asupra activității cerebrale sub influența psihedelicelor. Printre numeroasele lor utilizări, psihedelicele oferă și posibilități deosebite științei conștienței, deoarece alterează profund conținuturile conștienței ca urmare a simplei acțiuni farmacologice asupra creierului.

Chimistul elvețian Albert Hoffman, inventatorul dietilamidei acidului lisergic – LSD – arată cât de puternice pot fi aceste alterări într-o descriere a drumului său spre casă de la laboratoarele companiei farmaceutice Sandoz din Basel, pe 19 aprilie 1943. În acea zi, cunoscută acum ca „ziua bicicletei", a decis să înghită puțin din recenta sa descoperire. La scurt timp, a început să se simtă mai degrabă

* Introspecția reflectivă se păstrează în stările rare de „vis lucid", în care subiecții sunt conștienți de faptul că visează și își pot modifica voluntar comportamentul. Într-un studiu recent, absolut remarcabil, cercetătorii au putut comunica cu oameni care visau lucid. Aceștia se foloseau de mișcările ochilor, la fel ca pacienții locked-in descriși mai devreme, și au putut răspunde corect la întrebări simple de matematică și la alte întrebări cu răspuns de tip da/nu.

Scanned with OKEN Scanner

neobișnuit, așa că a plecat spre casă, pe bicicletă. După ce a reușit cumva să ajungă, deși a trecut prin tot felul de experiențe neplăcute, care l-au făcut să creadă că înnebunește, s-a întins pe canapea și a închis ochii.

...Puțin câte puțin am putut să încep să mă bucur de culorile nemaipomenite și de jocurile de forme care persistau în spatele ochilor mei închiși. Imagini caleidoscopice, fantastice năvăleau peste mine alternând, pestrițe, deschizându-se și închizându-se în cercuri și spirale, explodând în fântâni colorate, rearanjându-se și hibridizând-se într-un flux continuu...

În starea psihedelică, halucinații perceptuale realiste sunt însoțite adesea de experiențe neobișnuite despre sine, descrise îndeobște ca „disoluție a ego-ului", în care hotarele dintre sine și lume și alte persoane par să se deplaseze sau să se dizolve. Aceste evadări din experiența conștientă „normală" sunt atât de obișnuite, încât starea psihedelică ar putea reprezenta nu numai o schimbare în *conținutul* de conștiență, cât și în *nivelul* general de conștiență. Aceasta este ideea pe care am vrut să o testăm, în colaborare cu Robin Carhart-Harris, la Imperial College din Londra, și cu Suresh Muthukumaraswamy, la Universitatea din Auckland.

În aprilie 2016, Robin și cu mine am fost la o conferință la poalele dealurilor din Santa Catalina Mountains, în afara Tucsonului, Arizona. Fuseserăm amândoi invitați să ținem prelegeri despre cercetările noastre, astfel că ne-am folosit de împrejurare pentru a afla cum anume interesul fiecăruia pentru subiectul conștienței se suprapune cu al celuilalt. După zeci de ani în „sălbăticie", abia recent s-au reluat cercetările științifice și medicale asupra LSD și a altor compuși psihedelici precum psilocibina (ingredientul activ din ciupercile „magice"). După experimentul pe care

Scanned with OKEN Scanner

Hofmann l-a făcut asupra sa, a urmat o scurtă perioadă de efervescență pe această temă, mai multe studii ocupându-se de potențialul LSD de a trata un spectru de tulburări psihologice, inclusiv dependențele și alcoolismul, cu rezultate foarte promițătoare. Numai că folosirea, tot atunci, a LSD ca drog recreațional și ca simbol al rebeliunii, evanghelizate, printre alții, de Timothy Leary, a făcut ca aceste cercetări să fie stopate la finalul anilor 60. A fost nevoie să treacă zeci de ani până când, prin anii 2000, să se reia cercetarea serioasă – practic, s-a pierdut o generație de progres științific.

La nivel de neurochimie, psihedelicele clasice – LSD, psilocibina, mescalina și dimetiltriptamina (DMT, ingredientul activ din ayahuasca, băutura sud-americană halucinogenă) – acționează, în principal, prin influențarea sistemului de serotonină al creierului. Serotonina este unul dintre principalii neurotransmițători la nivel cerebral – substanțe care scaldă rețelele creierului și care influențează modul în care comunică neuronii. Psihedelicele influențează sistemul de serotonină prin legarea la un anumit receptor de serotonină, numit 5-HT_{2A}, care se găsește peste tot în creier. Una dintre principalele provocări în ce privește cercetarea psihedelicelor este de a înțelege cum anume aceste intervenții farmacologice minore reușesc să influențeze tiparul global al activității cerebrale, într-o asemenea măsură încât determină modificări profunde ale experienței conștiente.

Echipa lui Robin a descoperit anterior că starea indusă de psihedelice implică alterări surprinzătoare în dinamica cerebrală, comparativ cu martorul căruia i s-a administrat placebo. Rețelele cerebrale a căror activitate este corelată, de obicei – așa numitele rețele „de repaus" – se decuplează una de cealaltă, alte regiuni, de obicei mai mult sau mai puțin independente, cuplându-se între ele. Tabloul general descrie, astfel, o discontinuitate în tiparul de conectivitate care caracterizează activitatea cerebrală în condiții normale. Ideea lui Robin este că aceste discontinuități ar

Scanned with OKEN Scanner

fi responsabile de caracteristicile particulare ale stării psihedelice, cum ar fi disoluția granițelor dintre sine și lume sau amestecul simțurilor.

Robin și cu mine am dedus că datele pe care le-a colectat se potriveau perfect pentru analizele complexității algoritmice pe care le practicam cu echipa mea la Sussex, pe somn și anestezie. Ocazional, unele dintre scanările lor cerebrale s-au făcut cu magnetoencefalografie (MEG), care oferă rezoluția temporală mare și acoperirea cerebrală generală de care aveam nevoie. Au folosit MEG pentru a măsura activitatea cerebrală la voluntari care și-au administrat fie psilocibină, fie LSD, fie doze mici de ketamină. (În timp ce dozele mari de ketamină acționează ca anestezic, dozele mici au, mai degrabă, efect halucinogen.) Puteam folosi aceste date pentru a răspunde următoarei întrebări: ce se întâmplă cu valorile nivelurilor de conștiență, atunci când conținutul de conștiență se schimbă atât de dramatic în cursul călătoriei psihedelice?

La întoarcerea în Sussex, Michael Schartner și Adam Barrett au calculat schimbările din complexitatea algoritmică a semnalului MEG survenite în diverse regiuni cerebrale pentru toate cele trei stări psihedelice. Rezultatele au fost clare și surprinzătoare: psilocibina, LSD și ketamina au determinat *amplificări*, prin raportare la martorul placebo. A fost pentru prima dată când cineva a observat o creștere în conținutul de conștiență, comparativ cu nivelul de bază caracteristic stării de veghe în repaus. Toate comparațiile anterioare, fie pentru somn, anestezie sau tulburări de conștiență, au constatat descreșteri ale acestor valori.

Pentru a înțelege ce anume înseamnă aceste rezultate, precizez că măsurarea pe care o practicăm, cea a complexității algoritmice, reprezintă soluția cea mai bună în depistarea gradului de aleator sau a „diversității semnalului", în cazul semnalelor cerebrale pe care le analizează. O secvență complet aleatorie va avea cea mai înaltă complexitate algoritmică

Scanned with OKEN Scanner

posibilă, cea mai mare diversitate. Rezultatele noastre au venit, astfel, în completarea studiilor anterioare ale lui Robin, arătând că activitatea cerebrală în stadiul psihedelic devine tot mai aleatorie în timp, în linie cu reorganizarea detașată a experienței perceptuale pe care o descriu frecvent cei aflați într-o astfel de călătorie. De asemenea, aceste rezultate aduc lămuriri despre modul în care se corelează nivelul cu conținutul de conștiență – deoarece reprezintă un exemplu în care valoarea *nivelului* de conștiență se modifică din cauza schimbărilor generale ale *conținutului* de conștiență, schimbări ce caracterizează starea psihedelică. Faptul că valoarea nivelului de conștiență se modifică odată cu schimbarea valorii conținutului de conștiență atestă că cele două nu sunt aspecte independente ale conștienței.

Rezultatele analizelor noastre asupra psihedelicelor ridică, însă, o problemă tulburătoare. Oare gradul maxim de aleator din activitatea cerebrală ar putea determina un maxim în experiența psihedelică? Sau un „nivel" de conștiență diferit, de un alt tip? Extrapolarea nu pare aplicabilă. Un creier ai cărui neuroni sunt străbătuți de impulsuri electrice la întâmplare pare să nu genereze experiențe conștiente de niciun fel, așa cum formele libere de jazz nu mai sunt, la un moment dat, muzică.

Din păcate, complexitatea algoritmică este o aproximare slabă a ceea ce înseamnă, de obicei, „a fi complex". La prima vedere, complexitatea nu e același lucru cu aleatoriul. O noțiune mai potrivită complexității este precum calea de mijloc dintre ordine și dezordine – nu punctul maxim al dezordinii. Sunt Nina Simone și Thelonious Monk, nu trupa Bonzo Dog Doo-Dah.* Ce se întâmplă dacă abordăm, referitor la complexitate, acest mod de gândire mai sofisticat?

* În albumul lor de debut din 1967, numit Gorilla, trupa Bonzo Dog Doo-Dah au parodiat jazz-ul tradițional, interpretându-l cât mai prost posibil. Vom reveni la gorile în această carte.

Scanned with OKEN Scanner

Un articol din revista *Science*, publicat în 1998 de Giulio Tononi și de fostul meu șef și mentor, Gerald Edelman, face exact asta. Încă-mi amintesc de acest articol pe care l-am citit cu 20 de ani în urmă. A fost un eveniment care mi-a influențat decisiv gândirea despre conștiență, precum și un motiv esențial pentru care am ales să lucrez la Institutul de Neuroștiințe din San Diego.

În loc să se concentreze asupra unui singur caz clar de experiență conștientă, în stilul abordării „corelatelor neuronale ale conștienței" (NCC = neural correlates of consciousness) – cum este, de exemplu, experiența „vederii roșului" – Tononi și Edelman s-au întrebat care este caracteristica experienței conștiente în general. Ei au făcut o observație simplă, dar profundă: experiențele conștiente – *toate* experiențele conștiente – sunt atât *informative*, cât și *integrate*. Pornind de la această constatare, ei au făcut afirmații despre baza neuronală a fiecărei experiențe conștiente, nu numai a unei experiențe specifice cum ar fi vederea roșului, sentimentul de gelozie sau durerea de dinți.

Ideea de conștiență ca fiind simultan informativă și integrată necesită lămuriri suplimentare.

Să începem cu informația. Ce înseamnă când spui că experiențele conștiente sunt „informative"? Edelman și Tononi nu s-au referit la asta în același sens în care cititul unui ziar poate fi informativ, ci în sensul că, deși ar putea să pară banal, ascunde multă substanță. Experiențele conștiente sunt informative deoarece fiecare dintre ele este diferită de oricare altă experiență conștientă pe care ai avut-o, o vei avea sau ai fi putut să o ai vreodată.

Privind dincolo de biroul din fața mea, prin fereastră, nu am trăit niciodată *această* configurație alcătuită din cești le de cafea, monitoarele computerelor, nori – experiență care

Scanned with OKEN Scanner

este și mai individuală atunci când se combină cu toate celelalte percepții, emoții și gânduri prezente în fundalul universului meu interior. În fiecare secundă trăim *o* anume experiență conștientă din *infinitatea de posibile* experiențe conștiente. Fiecare experiență conștientă vine, astfel, cu o imensă reducere a incertitudinii, cât timp *ea* este cea trăită, nu o *alta* sau *cealaltă* șamd. Iar reducerea incertitudinii este – matematic – ceea ce se înțelege prin informație.

Capacitatea de a informa a unei anumite experiențe conștiente nu este o funcție de cât de bogată, detaliată sau de cât de revelatoare este pentru persoana care are respectiva experiență. Dacă asculți Nina Simone în timp ce mănânci căpșune într-un rollercoaster nu înseamnă că ai mai multe experiențe conștiente decât dacă ai asculta-o într-o cameră izolată, cu ochii închiși. Fiecare experiență reduce incertitudinea în raport cu numărul de experiențe posibile fix cu același număr.

Din această perspectivă, „cum ar fi să"-ul oricărei experiențe conștiente este definit nu atât prin ce anume este, ci prin toate cele nefăcute, dar posibile, prin care *nu* este. O experiență a roșului pur este așa cum este nu din cauza vreunei proprietăți intrinseci a „roșeții", ci pentru că roșul nu este albastru, verde sau orice altă culoare, sau orice miros, sau gând, sau sentiment de regret, sau orice altă formă de conținut mental, indiferent de care ar fi aceea. Roșul e roșu din cauza a tot ceea ce roșul nu este – la fel și pentru orice altă experiență conștientă.

Valorile mari la categoria informație nu sunt suficiente în sine. Experiențele conștiente nu sunt doar foarte informative, ci și *integrate*. Ce anume înseamnă atunci când afirmăm despre conștiență că este „integrată" reprezintă, încă, un subiect de dezbatere, dar înțelesul primar este că fiecare experiență conștientă pare a fi o perspectivă unificată. Nu experimentăm culorile separat de forma lor, nici obiectele separat de cadrul în care se

Scanned with OKEN Scanner

află. Nenumăratele elemente diferite din experiența mea conștientă din acest moment – computere și cești de cafea, sunetul ușii închizându-se undeva în hol și gândurile mele despre ce urmează să scriu – par întrețesute într-o manieră fundamentală și de neevitat, ca aspecte ale unei singure scene conștiente atot-cuprinzătoare.

Tononi și Edelman au venit cu o schimbare decisivă, propunând ca, dacă fiecare experiență conștientă este atât informativă, cât și unificată la nivel de fenomenologie, *atunci mecanismele neuronale aflate la baza experienței conștiente manifestă și ele ambele proprietăți.* Aceasta în virtutea faptului că exprimarea ambelor proprietăți – cu care mecanismele neuronale nu se corelează, ci de care chiar răspund – concentrează caracteristicile fenomenologice ale fiecărei experiențe conștiente.

Ce anume înseamnă, pentru un mecanism, să fie atât integrat, cât și informativ? Să abstractizăm un pic, dincolo de creier, și să considerăm un sistem alcătuit dintr-un număr mare de elemente care interacționează între ele, fără să ținem cont de ce anume sunt aceste elemente. Așa cum se observă în ilustrația de pe pagina următoare, pentru oricare astfel de sistem putem defini o scală cu două capete. La un capăt (cel din stânga) toate elementele se comportă aleatoriu și independent, precum moleculele unui gaz. Acest tip de sistem deține maximum de informație – aleatoriul este maxim – dar fără niciun fel de integrare, deoarece fiecare element este independent de toate celelalte.

La celălalt capăt (din dreapta), toate elementele fac exact același lucru, astfel că starea fiecărui element este complet determinată de starea celorlalte elemente din sistem. Nimic nu este aleatoriu. Ne putem gândi la aranjarea atomilor într-o rețea cristalină, unde poziția fiecărui atom este determinată complet de structura rețelei, care este definită de pozițiile tuturor celorlalți atomi. Acest tip de aranjare exprimă maximum de integrare, dar

Scanned with OKEN Scanner

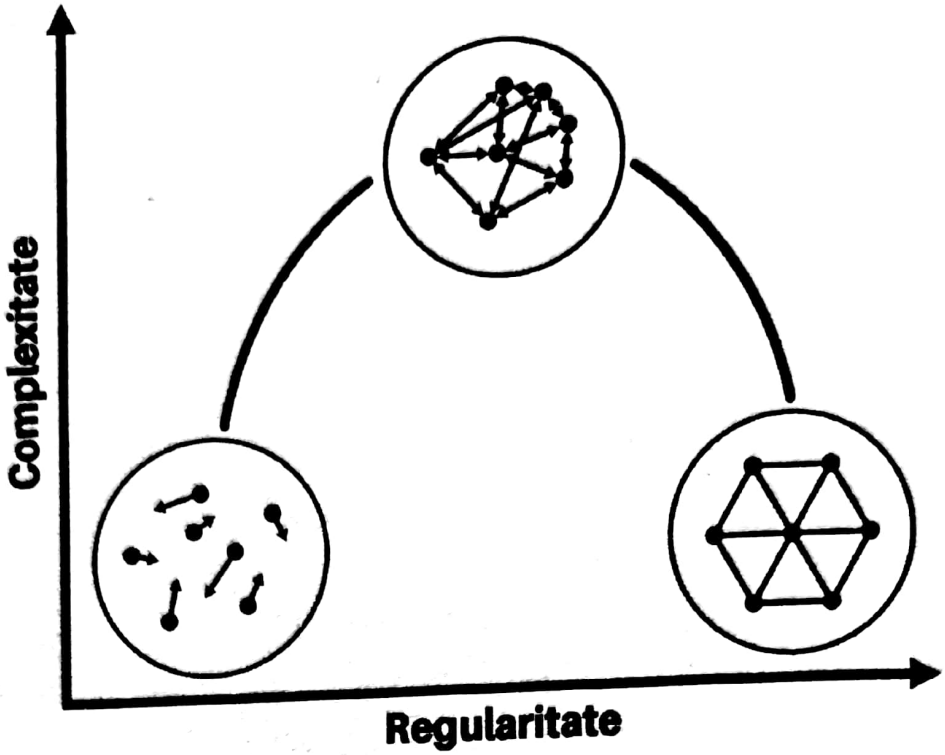

Fig. 2: Relația dintre complexitate și regularitate.

informație aproape zero, deoarece stările în care poate să existe sistemul sunt extrem de puține.

În mijloc se află sisteme ale căror elemente pot face diferite lucruri, dar în care există un anumit grad de coordonare, astfel că sistemul se comportă, într-o anumită măsură, ca „un întreg". Acesta este domeniul în care pot fi identificate atât informația, cât și integrarea. Este, de asemenea, domeniul de mijloc dintre ordine și dezordine și aici sunt situate sistemele numite, îndeobște, „complexe".

Atunci când aplicăm aceste descrieri creierului, avem o perspectivă mai clară asupra bazei neurale a conștienței.

Într-un creier cu informația la maximum, toți neuronii s-ar comporta independent, generând impulsuri electrice ca și cum ar fi complet deconectați. Într-un astfel de creier, complexitatea algoritmică (cum e complexitatea LZW), ar fi foarte mare. Dar acest creier – plin de informație, dar fără integrare – nu poate genera nicio stare conștientă. La extrema cealaltă, neuronii unui creier ordonat perfect ar face, cu toții, același lucru, probabil generând impulsuri

Scanned with OKEN Scanner

în același ritm, cam ca în situația crizelor epileptice. Complexitatea algoritmică ar fi, aici, extrem de mică. Și acest tip de creier ar fi lipsit de conștiență, dar dintr-un cu totul alt motiv: foarte multă integrare, informație zero.

O măsură potrivită a nivelului de conștiență ar trebui, deci, să nu urmărească informația în sine, ci cum anume se exprimă împreună informația și integrarea. O astfel de măsură – a complexității în adevăratul sens al cuvântului – ar exemplifica abordarea conștienței, de către adevărata problemă, legând explicit proprietățile mecanismului de proprietățile experienței.

Așa cum am văzut, aproximările făcute prin complexitatea algoritmică, cum este LZW, nu sunt prea grozave din acest punct de vedere. Ne pot oferi multe date referitor la informație, dar nimic despre integrare. PCI se descurcă ceva mai bine. Pentru o valoare mare pe scala PCI, pulsul de energie injectată de TMS trebuie să genereze un tipar al activității cerebrale care să fie dificil de comprimat, deci plin de informație. Pulsul trebuie și să se răspândească departe pe toate direcțiile cortexului, pentru a genera un „ecou" – a cărui compresibilitate să poată fi, apoi, evaluată. Totuși, deși această răspândire corticală este o expresie a integrării, tot nu este suficientă pentru ceea ce avem nevoie, la modul ideal, de la o astfel de măsurătoare. Măsurarea PCI depinde de integrarea activității cerebrale într-o manieră mai degrabă vagă – altfel, nu s-ar genera ecou – dar nu măsoară integrarea în aceeași manieră cantitativă în care este măsurată informația. Ceea ce căutăm sunt măsurători care reflectă atât integrarea, cât și informația, pornind de la aceleași date, în același timp.

Există câteva măsurători care îndeplinesc aceste criterii, măcar în teorie. În anii 90, Tononi și Edelman, împreună cu Olaf Sporns, colegul lor, au sugerat o măsură pe care au numit-o „complexitate neuronală", iar 10 ani mai târziu am continuat ideea sugerând propria măsură, pe care

Scanned with OKEN Scanner

am numit-o „densitate cauzală", folosind un alt tip de matematici. Mai multe măsurători noi – unele pe care le vom întâlni în capitolele următoare – au fost generate pe căi incredibil de sofisticate, pornind de pe aceste baze. Toate aceste măsurători intenționează să cuantifice, într-un fel sau altul, cât anume ocupă un sistem din zona aceea de mijloc dintre ordine și dezordine, zonă în care se poate găsi integrarea. Problema e că, totuși, niciuna nu a funcționat cum trebuie, atunci când a fost aplicată datelor obținute prin neuroimagistică.

E ceva curios în legătură cu această situație. Ne puteam aștepta ca măsurătorile care reflectă îndeaproape principii teoretice să fie mai bune, în practică, decât măsurătorile precum complexitatea algoritmică, foarte puțin conectate cu o anume teorie. Numai că nu asta constatăm, deci ce se întâmplă, oare? O posibilitate este ca teoria în sine să fie eronată. Totuși, eu cred că trebuie doar să lucrăm mai mult la rafinarea aparatului matematic, astfel încât măsurătorile să facă ceea ce dorim noi să facă. De asemenea, ar fi nevoie de dezvoltarea de noi metode de imagistică cerebrală pentru a obține acele date pe care să le putem interpreta.

Căutarea unui conștientometru adevărat continuă, deci. E important să evidențiem progresul de până acum, care e remarcabil. Este recunoscut faptul că nivelul de conștiență nu este totuna cu starea de veghe și avem deja mai multe măsurători ale nivelului de conștiență, pe baza analizei creierului, măsurători care ajută enorm la depistarea diferitelor stări globale de conștiență și în detectarea conștienței reziduale la pacienții cu leziuni cerebrale. PCI fundamentat de Massimini este chiar semnificativ. Util clinic și bine fundamentat pe principii teoretice clar formulate despre informație și integrare,

Scanned with OKEN Scanner

conectează eficient mecanismele neuronale cu proprietățile universale ale experienței conștiente – în stilul „problemei reale". Apar mereu și alte măsurători, bazate pe principii diferite, dar înrudite, iar aproximările ușor de folosit, cum ar fi estimarea complexității algoritmice a datelor cerebrale spontane, relevă conexiuni fascinante între nivelul de conștiență și conținutul conștienței.

Rămâne, totuși, o întrebare fundamentală. Este conștiența mai degrabă ca *temperatura* – reductibilă la și identificabilă cu o proprietate fundamentală a universului fizic (sau informațional)? Abordările privind măsurarea conștienței pe care le-am întâlnit până acum se bazează pe povestea temperaturii, dar eu cred că, în final, s-ar potrivi mai bine cu o analogie din *viață*. Pentru mine, „integrare" și „informație" sunt proprietăți generale ale celor mai multe – dacă nu toate – experiențe conștiente. Asta nu înseamnă că, dacă temperatura *este* media energiei cinetice moleculare, atunci conștiența *este* informație integrată.

Pentru a vedea ce anume mă face să cred astfel, trebuie să împingem analogia dintre conștiență și temperatură cât de departe se poate, pentru a vedea dacă și când va ceda. Este timpul să vă prezint „teoria informației integrate" a conștienței.

Scanned with OKEN Scanner

3
Phi

Este iulie 2006 și sunt în Las Vegas, la o înghețată cu Giulio Tononi. Suntem la hotelul Venețian și nu prea știu ce se întâmplă. Am zburat cu o zi înainte de la Londra, suntem undeva în interiorul Venețianului, într-o încăpere în care se lasă mereu seara, stelele false tocmai încep să sclipească pe cerul fals azuriu, în timp ce gondolele false plutesc pe lângă palatele false. Asta pentru ca oamenii să-și cheltuie acolo banii, într-o permanentă stare de *aperitivo*, fără să conștientizeze cât timp a trecut. Sufăr din cauza decalajului de fus orar, sunt ușor amețit – ne-am mutat aici după o cină lungă – și dezbatem, de câteva ore, detaliile ambițioasei „teorii a informației integrate" a conștienței – sau IIT (integrated information theory). Este creația minții lui Tononi și, mai mult decât orice altă teorie susținută neuroștiințific, abordează din prima problema grea a conștienței. IIT afirmă că experiența subiectivă este o proprietate a tiparelor cauzei și efectului, că informația este la fel de reală ca masa sau ca energia și că chiar și atomii ar putea fi un pic conștienți.

Nu este o discuție echitabilă. Îmi petrec mult timp argumentând o lucrare recentă de-a mea în care critic o versiune timpurie a teoriei sale. Giulio încearcă să-mi explice de ce greșesc, blând, dar insistent. Nu sunt sigur ce anume – decalajul de fus orar, vinul sau logica necruțătoare a lui Giulio – mă face să mă simt mai puțin sigur pe mine decât eram în timpul zborului. A doua zi dimineața mă hotărăsc să mă gândesc mai profund, să înțeleg mai mult, să mă pregătesc mai bine și să beau mai puțin.

Scanned with OKEN Scanner

IIT mi s-a părut fascinantă atunci și mi se pare la fel și acum, pentru că exemplifică analogia dintre conștiență și temperatură. Potrivit IIT, conștiența *este*, pur și simplu, informație integrată. Raționamentul teoriei se bazează pe respingerea opiniilor vehiculate până atunci despre maniera în care mintea și materia se interconectează și despre cum se împletește conștiența în structura universului.

Pe atunci, în 2006, IIT nu era bine cunoscută. Astăzi, este una dintre cele mai bine cotate, dar și mai aprins dezbătute teorii din știința conștienței. A fost elogiată de unele dintre cele mai mari nume din domeniu, în afară de Tononi însuși. Christof Koch, fost susținător al abordării NCC, a numit-o „un pas gigantic în rezolvarea finală a vechii probleme minte-corp". Dar aspirațiile și importanța sa au provocat și o respingere considerabilă. Unul dintre motive este abordarea profund matematică și mult prea complexă. Desigur, acesta nu este neapărat un lucru rău: nimeni nu spune că rezolvarea problemei conștienței ar trebui să fie ușoară. O altă obiecție este că afirmațiile pe care le face sunt atât de contraintuitive, încât teoria trebuie că este greșită. Și aceasta este o opinie periculoasă atunci când te confrunți cu un fenomen atât de șocant cum este conștiența.

Pentru mine, problema mai grea este că extraordinarele afirmații ale IIT necesită dovezi extraordinare, iar fix ceea ce-și dorește IIT – adică să rezolve problema grea – face ca afirmațiile sale să fie de netestat în practică. Dovezile extraordinare necesare nu pot fi obținute. Din fericire, nu este totul pierdut. După cum voi explica, unele predicții ale IIT pot fi testabile, cel puțin în principiu. Și există interpretări alternative ale IIT, mai apropiate de problema reală decât de problema grea, care conduc la dezvoltarea de noi măsurători ale nivelului conștienței, fundamentate teoretic, *dar și* aplicabile practic.

Așa cum dezvăluie chiar numele teoriei, conceptele de „informație" și „integrare" reprezintă esența IIT. Teoria se

Scanned with OKEN Scanner

bazează pe ideile despre măsurarea nivelului conștienței pe care le-am întâlnit și în capitolul anterior, numai că în maniere total diferite.

În inima IIT se află o singură mărime numită „Φ" (litera greacă phi, pronunțată *fi*). Cel mai simplu mod de a înțelege Φ este de a te gândi că măsoară cât de mult este un sistem „mai mult decât suma" părților sale, în termeni de informație. Cum poate fi un sistem mai mult decât suma părților sale? Ne putem gândi, prin analogie, la un stol de păsări: stolul pare să fie mai mult decât suma păsărilor care îl alcătuiesc – pare să aibă o „viață a sa". IIT preia această idee și o transpune în domeniul informației. În IIT, Φ măsoară cantitatea de informație pe care un sistem o generează „ca întreg", dincolo de cantitatea de informație generată de fiecare dintre părțile sale. Pe asta se bazează principala afirmație a IIT cum că *un sistem este conștient în măsura în care generează, ca întreg, mai multă informație decât părțile sale.*

De observat că nu este vorba nici despre o propunere referitoare la o corelație și nici despre o propunere de tip adevărata-problemă în ceea ce privește modul în care proprietățile mecaniciste ale unui sistem răspund de proprietățile fenomenologiei. Este o propunere referitoare la identitate. Conform IIT, nivelul Φ este intrinsec unui sistem (adică nu depinde de un observator extern) și este identic cu cantitatea de conștiență asociată cu acel sistem. Cu cât Φ este mai ridicat, cu atât mai multă conștiență. Zero Φ, zero conștiență. De aceea, IIT este expresia supremă a unei viziuni despre conștiență similară cu măsurarea temperaturii.

Cum facem să avem Φ mare? Deși ideea de bază ar trebui să fie cunoscută din capitolul anterior, există câteva diferențe importante, așa că merită să o discutăm de la zero.

Imaginați-vă o rețea de „neuroni" artificiali simplificați, fiecare dintre ei putând fi „activat" (on) sau „dezactivat"

Scanned with OKEN Scanner

(off). Pentru un Φ mare, rețeaua trebuie să îndeplinească două condiții principale. În primul rând, starea generală a rețelei – rețeaua „ca întreg" – trebuie să elimine un mare număr de alternative posibile la rețeaua globală. Aceasta este *informație* și reflectă observația din fenomenologie conform căreia fiecare experiență conștientă elimină o mulțime de alte experiențe conștiente posibile. În al doilea rând, trebuie să extragem mai multă informație de la sistemul luat ca întreg decât de la părțile sale (neuronii individuali sau grupurile de neuroni) luate fiecare separat. Aceasta este *integrarea* și reflectă observația că toate experiențele conștiente sunt aduse împreună, că sunt experimentate „ca întreg". Φ este o modalitate de a asocia un număr unui sistem care măsoară cât de mari sunt valorile ambelor dimensiuni.

Există multe moduri în care un sistem poate eșua în obținerea unui Φ ridicat. Unul este obținerea unui scor mic la informație. Un exemplu banal este o fotodiodă – un senzor simplu de lumină care poate fi ori „pornit", ori „oprit". În acest caz, Φ este scăzut sau zero, deoarece starea sa, în orice moment, este caracterizată de foarte puțină informație despre orice. Indiferent de starea în care se află (unu sau zero, pornit sau oprit), rezultă întotdeauna o singură alternativă (zero sau unu). O fotodiodă transmite maximum un „bit" de informație.*

Un alt mod în care Φ rămâne scăzut este un scor mic la integrare. Să ne imaginăm un câmp de fotodiode, ca senzorul din camera telefonului, de exemplu. Starea globală a sistemului este starea întregului câmp, care poate transporta multă informație. Un câmp suficient de mare de senzori va intra într-o stare globală – diferită pentru fiecare stare a lumii pe care o întâlnește, motiv pentru care camerele sunt utile. Dar această informație globală nu contează pentru senzorul însuși. Fiecare fotodiodă

* În teoria informației, „bitul" este unitatea fundamentală de informație.

Scanned with OKEN Scanner

din senzor este independentă cauzal de celelalte – starea fiecăreia depinde doar de nivelul de lumină pe care îl întâlnește. Tăiați senzorul într-o grămadă de fotodiode (independente din punct de vedere cauzal) și veți vedea că va funcționa la fel de bine. Informația transmisă de câmpul de senzori în ansamblu nu este mai mult decât informația transmisă independent de fiecare senzor – de toate fotodiodele – însă. Informația pe care o generează nu este mai mult decât suma părților sale și, astfel, Φ va fi, de asemenea, zero.

Un alt exemplu instructiv de Φ zero este situația cunoscută sub denumirea de „creier divizat". Să ne imaginăm o rețea împărțită în două jumătăți complet separate. Fiecare jumătate a acestei rețele poate avea un Φ diferit de zero, dar rețeaua ca întreg va avea Φ zero. Aceasta deoarece există o modalitate de a împărți rețeaua în părți – cele două jumătăți – pentru care întregul nu este mai mult decât suma părților. Acest exemplu subliniază modul în care Φ depinde de modul optim în care un sistem poate fi „tăiat", pentru a minimiza diferența dintre ceea ce face întregul și ceea ce fac părțile. Acesta este unul dintre aspectele distinctive ale IIT, care îl diferențiază de măsurătorile de complexitate descrise în capitolul anterior.

Acest exemplu implică, de asemenea, că separarea *reală* a unui creier în cele două emisfere – în urma unei intervenții chirurgicale care se practică în cazuri de epilepsie, altfel netratabilă – ar putea face ca un creier să aibă două „conștiențe" independente, alocate, fiecare, câte unei emisfere cerebrale, pentru că nu va exista o singură conștiență care să aparțină ambelor emisfere. Tot așa, și tu și eu suntem amândoi conștienți, dar nu există o conștiență colectivă care să aparțină amândurora, deoarece putem fi divizați informațional între noi.

În ce privește organul real, creierul, IIT susține cu argumente clare o serie de observații despre nivelul de

Scanned with OKEN Scanner

conștiență. Știm din capitolul anterior că cerebelul nu prea pare implicat în conștiență, deși conține aproximativ trei sferturi din toți neuronii din creier. Acest lucru este explicat de IIT prin faptul că anatomia cerebelului este comparabilă cu matricea de senzori din camera foto a telefonului – un număr mare de circuite semi-independente neprietenoase pentru Φ. Cortexul cerebral, pe de altă parte, este plin de rețele extrem de interconectate, ceea ce se poate asocia cu Φ ridicat. Atunci de ce se estompează conștiența în timpul somnului fără vise, a anesteziei și a comei, dacă aceste circuite nu se modifică? IIT stipulează că, în aceste stări, capacitatea neuronilor corticali de a interacționa unii cu alții este atât de compromisă încât Φ dispare.

IIT este o abordare „axiomatică" a conștienței. Începe mai degrabă cu principii teoretice, decât cu date experimentale. În logică, axiomele sunt afirmații evident adevărate, în sensul că se admite despre ele, la modul general, că nu necesită explicații suplimentare. „Două forme care umplu perfect același spațiu sunt aceeași formă" – cum spunea filozoful grec Euclid – este un bun exemplu. IIT propune axiome despre conștiență – în primul rând, experiențele conștiente sunt atât integrate, cât și informative – și utilizează aceste axiome pentru a susține afirmațiile despre ce proprietăți trebuie să aibă mecanismele care stau la baza acestor experiențe. În IIT, *orice* mecanism care are aceste proprietăți, fie creier sau nu, biologic sau nu, va avea Φ diferit de zero și va avea conștiență.

Atât despre principii. Ca orice teorie, IIT va rezista sau nu în măsura în care predicțiile sale sunt testabile. Principala afirmație a teoriei este următoarea: nivelul de conștiență al unui sistem este dat de Φ. Testarea acestei afirmații necesită măsurarea Φ pentru sisteme reale, iar

Scanned with OKEN Scanner

de aici încep problemele. Se pare că măsurarea Φ este extrem de dificilă și, în majoritatea cazurilor, aproape, sau, de fapt, imposibilă. Motivul principal este că IIT tratează „informația" într-un mod neobișnuit.

Utilizarea standard a informației în matematică, dezvoltată de Claude Shannon în anii 1950, este *legată de observator*. Informația legată de observator (sau *extrinsecă*) reprezintă măsura în care a fost redusă incertitudinea, din perspectiva unui observator care urmărește un sistem într-o anumită stare. Cum ar fi, de exemplu, când arunci un singur zar de mai multe ori. De fiecare dată, observi un singur rezultat din șase posibilități – celelalte cinci alternative sunt excluse. Asta corespunde cu reducerea incertitudinii unei anumite cantități (măsurată în biți) și reprezintă o informație care este „pentru" observator.

Pentru a măsura informația legată de observator, de obicei este suficient să vezi cum se comportă un sistem, pe o anumită perioadă de timp. Cu zarurile, poți, pur și simplu, să consemnezi ceea ce obții la fiecare aruncare, ceea ce-ți permite să calculezi câtă informație este generată prin aruncarea oricăruia dintre numere. Pentru un sistem alcătuit dintr-o rețea de neuroni, e suficient să înregistrezi activitatea neuronilor în timp. Un observator extern poate înregistra toate stările diferite în care intră neuronii, poate calcula probabilitățile asociate cu fiecare stare și apoi poate măsura reducerea incertitudinii asociată cu faptul că rețeaua se află în oricare dintre aceste stări.

Totuși, pentru IIT, informația nu poate fi tratată în acest mod legat-de-observator. Asta pentru că în IIT, informația – informație *integrată*, Φ – este, de fapt, conștiență și, deci, dacă tratăm informația ca fiind legată-de-observator, atunci și conștiența ar fi, la fel, legată-de-observator. Numai că nu este legată-de-observator. Că sunt sau că nu sunt conștient – asta nu ar trebui și nici nu depinde de modul în care cineva îmi măsoară creierul.

Scanned with OKEN Scanner

Prin urmare, în IIT, informația trebuie considerată ca *intrinsecă* unui sistem, nu ca fiind legată de un observator extern. Trebuie definită într-un mod care să nu depindă de niciun observator extern. Trebuie să fie informație „pentru" sistemul în sine – nu pentru oricine sau pentru orice altceva. Dacă nu, relația de identitate dintre Φ și conștiență din inima IIT nu poate rezista.

Pentru a măsura informația intrinsecă, nu este suficient doar să observi cum se comportă sistemul în timp. Tu – ca om de știință, observatorul extern – trebuie să cunoști toate modurile diferite în care un sistem *s-ar putea* comporta, chiar dacă, de fapt, nu se comportă niciodată în toate acele moduri. Diferența este între a ști *ce face, de fapt,* un sistem în timp (ceea ce e ușor, cel puțin în principiu, și e legat de observator) și a ști *ce ar putea face* un sistem, chiar dacă nu o face niciodată (ceea ce e, de obicei, dificil, dacă nu imposibil, dar este independent de observator).

În limbajul teoriei informației, diferența dintre aceste situații este aceeași ca dintre distribuția „empirică" a stărilor unui sistem și distribuția „entropiei lui maxime" (denumire care reflectă nivelul maxim de incertitudine cu privire la un sistem). Imaginează-ți că arunci două zaruri de mai multe ori. Poate arunci un șapte, un opt și un unsprezece și alte câteva numere, dar niciodată un doisprezece. În această situație, distribuția empirică nu ar conține un doisprezece, dar distribuția entropiei maxime l-ar conține, deoarece un doisprezece *s-ar fi putut întâmpla,* chiar dacă în această secvență anume de aruncări nu s-a întâmplat. Aceasta înseamnă că oricare rezultat – fie un șapte, un opt sau un unsprezece – ar genera mai multă informație (adică reduce mai mult incertitudinea, exclude mai multe alternative) despre distribuția entropiei maxime (care include doisprezece), decât despre distribuția empirică (care nu include doisprezece).

În comparație cu măsurarea distribuției empirice a unui sistem doar observându-l în timp, măsurarea distribuției

Scanned with OKEN Scanner

entropiei maxime este, în general, foarte dificil de făcut. Există două moduri în care am putea proceda. Primul constă în a perturba sistemul în toate modurile posibile și a vedea ce se întâmplă, la fel cum un copil ar putea apăsa toate butoanele unei jucării noi pentru a vedea toate lucrurile pe care le poate face. Al doilea constă în a deduce distribuția entropiei maxime din cunoașterea exhaustivă, completă a mecanismului fizic al sistemului – „structura sa cauză-efect". Dacă știi totul despre un mecanism, atunci este posibil, uneori, să știi toate lucrurile pe care le poate face, chiar dacă în practică nu le face. Dacă știu că un zar are șase fețe, îmi pot da seama că două zaruri pot genera toate numerele de la doi la doisprezece, fără să fie nevoie, însă, de nicio aruncare.

Din păcate, singurul lucru la care avem acces, de multe ori, este, mai degrabă, dinamica unui sistem, adică ceea ce *face* acel sistem, nu ceea ce *ar putea face*. Asta este, cu siguranță, valabil și pentru creier. Pot să înregistrez ceea ce face un creier, la diferite niveluri de detaliu, dar nu am de unde să-i cunosc complet structura fizică și nici nu-i pot perturba activitatea în toate modurile posibile. Din aceste motive, afirmația cea mai caracteristică a IIT – că Φ *este*, de fapt, conștiență – e și cea mai puțin testabilă.

Măsurarea Φ întâmpină și alte piedici, indiferent de tipul de informație vizat. Una este că măsurarea oricărui Φ impune împărțirea adecvată a sistemului, astfel încât să fie posibilă compararea „întregului" cu „părțile" sale. Pentru unele sisteme – cum ar fi emisferele cerebrale separate – este destul de ușor (împărțirea e la mijloc) dar, în general, aceasta este o problemă foarte dificilă, deoarece numărul de posibilități de împărțire a unui sistem crește exponențial cu dimensiunea acestuia.

Scanned with OKEN Scanner

Se profilează, apoi, și mai multe întrebări fundamentale despre ce anume poate fi considerat un sistem, în primul rând. Care este granularitatea corectă a spațiului și timpului pe baza căreia să se calculeze Φ? Vorbim de neuroni și milisecunde, sau de atomi și femtosecunde? Poate o țară întreagă să fie conștientă – poate o țară să fie mai conștientă decât alta? Am putea lua în considerare interacțiunile plăcilor tectonice pe scări de timp geologice ca integratoare de informații la scară planetară?

Este important să recunoaștem că aceste provocări – inclusiv cea de a măsura mai degrabă informația intrinsecă, decât informația extrinsecă, legată-de-observator – sunt probleme doar pentru noi, ca oameni de știință, ca observatori externi, care încercăm să calculăm Φ. Potrivit IIT, un sistem anume *ar avea* un Φ și atât. Și-ar vedea de treabă, integrând informația exact în același mod în care, atunci când arunci o piatră, aceasta trasează un arc prin aer fără a fi nevoie să-și calculeze traiectoria în conformitate cu legile gravitației. Doar pentru că o teorie este dificil de testat nu înseamnă că este și greșită. Înseamnă doar că este dificil de testat.

Dacă am lăsa la o parte provocarea de a măsura Φ, atunci ce anume ar implica IIT, dacă teoria este corectă? Mergând până la capăt pe calea IIT, ajungem la niște consecințe foarte ciudate.

Imaginează-ți că-ți deschid craniul și îți strecor în creier un pumn de neuroni noi, fiecare dintre ei conectându-se într-un anume fel la materia ta cenușie deja existentă. Mai mult, pe măsură ce îți desfășori activitatea zilnică, acești neuroni noi nu fac niciodată nimic. Indiferent de ce se întâmplă, indiferent de ceea ce faci sau pe cine vezi, nu declanșează niciodată impulsuri. Creierul tău îmbunătățit

Scanned with OKEN Scanner

pare, de fapt, fix la fel ca cel vechi. Noii tăi neuroni sunt, însă, organizați în așa fel încât *ar putea* declanșa impulsuri doar în situația în care restul creierului tău ar intra într-o anumită stare în care nu intră, de fapt, niciodată.

De exemplu, să presupunem că acești neuroni noi ar declanșa impulsuri doar dacă ar fi să mănânci un pepene Densuke, un fruct rar, care se găsește doar pe insula japoneză Hokkaido. Presupunând că nu mănânci niciodată un Densuke, deci acești noi neuroni nu declanșează niciodată impulsuri, IIT prezice, totuși, că *toate* experiențele tale conștiente s-ar schimba – deși foarte subtil. Asta pentru că acum există mai multe stări potențiale în care s-ar putea afla creierul tău – noii neuroni ar putea declanșa impulsuri – așa că și Φ trebuie să se schimbe.

Reversul acestui scenariu duce la o predicție la fel de ciudată. Imaginează-ți un grup de neuroni în repaus, situați în profunzimea cortexului vizual. Deși sunt conectați la alți neuroni – capabili, deci, să declanșeze impulsuri dacă sunt stimulați adecvat – acești neuroni nu fac nimic. Apoi, printr-o intervenție experimentală inteligentă, aceștia sunt împiedicați în mod activ să declanșeze impulsuri – devin mai degrabă *inactivați,* decât pur și simplu *inactivi.* Chiar dacă activitatea generală a creierului nu s-a schimbat deloc, IIT ar prezice din nou o schimbare în experiența conștientă, având în vedere că, acum, stările potențiale în care ar putea intra creierul sunt mai puține.

În mod remarcabil, o versiune a acestui experiment ar putea fi posibilă cât de curând, datorită noii tehnologii a optogeneticii, care permite cercetătorilor să controleze activitatea anumitor neuroni în detaliu. Optogenetica folosește tehnici genetice care modifică anumiți neuroni astfel încât aceștia să devină sensibili la lumină cu anumite lungimi de undă. Apoi, folosind lasere sau rețele de LED-uri pentru a proiecta lumină în creierul animalelor modificate genetic, acești neuroni pot fi „aprinși" sau

Scanned with OKEN Scanner

„stinși". În principiu, optogenetica ar putea fi folosită pentru *a inactiva* neuroni *deja inactivi*, permițând evaluarea efectelor asupra percepției conștiente – dacă există. Experimentul nu este unul simplu și nici nu oferă o modalitate de măsurare a Φ. Dar perspectiva de a testa orice aspect al IIT este incitantă și am avut norocul să mă implic în discuțiile recente – cu Giulio Tononi și alții – cu o perspectivă ce aduce o rezolvare.

Mărind perspectiva, o altă ciudățenie a IIT este că, afirmând cu convingere despre Φ că *este* conștiență, IIT implică, de asemenea, și că *informația în sine* există – că are o stare ontologică definită în universul nostru – o stare precum masa/energia și sarcina electrică. (Ontologia este studiul a „ceea ce există".) Într-un sens, este o aliniere cu viziunea „it din bit" a fizicianului John Wheeler, probabil cel mai cunoscut susținător al ideii conform căreia tot ceea ce există derivă, în final, din informație – că informația este originea și că orice altceva decurge din ea.

Iar asta duce la o concluzie finală provocatoare: *panpsihismul*. Cât timp există tipul potrivit de mecanism, tipul corect de structură cauză-efect într-un sistem, Φ va fi diferit de zero, deci va exista conștiență. Panpsihismul IIT este unul restrâns, nu coincide cu acela în care conștiența este răspândită peste întregul univers, ca un strat subțire de gem. Conștiența se găsește, mai degrabă, oriunde este informație integrată – Φ. Adică ici și acolo, dar nu peste tot.

IIT este originală, ambițioasă și exuberantă din punct de vedere intelectual. Rămâne singura teorie neuroștiințifică ce abordează serios problema grea a conștienței. IIT este și ciudată, cu siguranță, dar faptul că ceva este ciudat nu înseamnă că este și greșit. Aproape totul despre fizica modernă este ciudat, dar mai puțin greșit decât fizica

Scanned with OKEN Scanner

din trecut. Dar succesul acelor părți ale fizicii moderne, despre care știm acum că sunt mai puțin greșite, se datorează faptului că sunt testabile experimental. Și aceasta este problema cu IIT. Îndrăzneala ei vine cu prețul pe măsură pe care îl plătește: afirmația sa principală – echivalența dintre Φ și nivelul de conștiență – ar putea fi imposibil de testat.

Pentru posibilitățile actuale, cea mai bună soluție este să reținem ideea fundamentală a IIT, anume că experiențele conștiente sunt atât informative, cât și integrate, dar să renunțăm la pretenția că Φ ar fi, pentru conștiență, ceea ce este energia cinetică moleculară medie pentru temperatură. Astfel, perspectiva IIT despre structura experiențelor conștiente s-ar alinia cu cea a problemei reale. Adoptarea acestui punct de vedere deschide calea dezvoltării unor versiuni alternative, aplicabile practic, ale lui Φ, măsurători care ajung să aibă multe în comun cu măsurătorile complexității, întâlnite la finalul capitolului precedent.

Colegii mei, Adam Barrett și Pedro Mediano, și cu mine urmăm această strategie de mulți ani. Am dezvoltat mai multe versiuni ale lui Φ care funcționează mai degrabă cu informație legată-de-observator, decât cu informație intrinsecă. Aceasta ne permite să măsurăm Φ pe baza comportamentului observabil al unui sistem de-a lungul timpului, fără să ne îngrijorăm de ce anume ar putea face, dar nu o face niciodată. Așa cum stau lucrurile în prezent, versiunile noastre de Φ se comportă destul de diferit, chiar și pe modele de sistem foarte simple. Asta înseamnă că mai sunt încă multe de făcut în dezvoltarea versiunilor de Φ care funcționează în practică și că – sperăm – își câștigă greutatea empirică datorită, nu în ciuda, fundamentării lor prin principiile teoretice. Din perspectiva noastră, asta înseamnă să tratăm „integrarea“ și „informația“ ca proprietăți generale ale experiențelor conștiente ce trebuie explicate, nu ca pretenții axiomatice despre ceea ce „este“ de fapt conștiența. Cu alte cuvinte, să tratăm conștiența ca fiind mai asemănătoare cu viața decât cu temperatura.

Scanned with OKEN Scanner

Călătoria noastră prin nivelurile de conștiență ne-a scos din uitarea anesteziei și a comei, trecând prin zonele crepusculare ale stărilor vegetative și de minimă conștiență, prin lumile deconectate ale somnului și viselor, în lumina soarelui conștientizării depline a stării de veghe și chiar mai departe, spre ciudata hiperrealitate a psihedelicului. Toate aceste niveluri sunt conectate prin ideea că fiecare experiență conștientă este atât informativă, cât și integrată, situată într-o complexă zonă de mijloc dintre ordine și dezordine. Această idee fundamentală a dat naștere la noi măsurători, cum ar fi PCI, care sunt atât utile practic, cât și capabile să construiască punți explicative, în stilul adevărata-problemă, între fizic și fenomenal. Cu IIT, am ajuns la una dintre cele mai incitante și controversate frontiere ale științei conștienței, unde îndrăzneala întâlnește limitele testabilității și unde analogia dintre conștiență și temperatură poate ceda. Și, deși sceptic în ceea ce privește afirmațiile mai cuprinzătoare ale acestei teorii provocatoare, sunt la fel de dornic să văd cum se va dezvolta, ca acum mulți ani în urmă, când mâncam înghețată cu Giulio Tononi.

Privind în urmă, Las Vegas a fost locul cel mai potrivit pentru a dezbate IIT. Este informația reală? Este conștiența peste tot? În Las Vegas, e greu de crezut că ceva este real, în afară de senzația brută a experienței în sine. Chiar și acum, ani mai târziu, îmi pot imagina perpetuul amurg din hotelul Venețian, gondolele false urmându-și cu precizie traseele, ca niște mecanisme de ceas. Sunt, cu siguranță, conștient, dar *de ce anume* sunt conștient? În Venețian, este tentant să crezi că totul este un fel de halucinație.

Așa cum vom vedea, există un adevăr neașteptat în acest gând ciudat.

Scanned with OKEN Scanner

II

CONȚINUT

Scanned with OKEN Scanner

4

Percepția în sens invers

Deschid ochii și îmi apare imaginea unei lumi. Stau pe veranda unei case din lemn, așezată sus, pe versantul unei păduri de chiparoși, la câțiva km nord de Santa Cruz, California. E dimineața devreme. Privind în fața mea, pot vedea arbori înalți, încă învăluiți de ceața rece a oceanului, adunată noapte de noapte, scăzând mult temperatura. Nu pot vedea solul, așa că terasa și arborii par să plutească, împreună cu mine, în abur. Sunt câteva scaune vechi de plastic – stau pe unul –, o masă, o tavă cu pâine și cafea. Aud cântecele păsărilor, zgomote în spatele meu – oamenii din casă – și un murmur îndepărtat, de la ceva ce nu pot identifica. Nu toate diminețile sunt astfel; aceasta este o dimineață frumoasă. Încerc să mă conving pe mine însumi, a nu știu câta oară, că această lume extraordinară este o construcție a propriului meu creier, un fel de „halucinație controlată".

Ori de câte ori suntem conștienți, suntem conștienți de ceva sau de mai multe lucruri. Acestea sunt *conținutul* conștienței. Pentru a înțelege cum anume apare și ce anume vreau să spun prin halucinație controlată, să schimbăm perspectiva. Imaginează-ți, pentru o secundă, că ești un creier.

Adică străduiește-te să te gândești cum ar fi să fii aici sus, încuiat în bolta osoasă a craniului, încercând să-ți dai seama ce e acolo, în lumea de afară. Nu e lumină, nici sunet, nimic – e complet întuneric și extrem de liniște. Atunci când încearcă să alcătuiască percepții, tot ceea ce trebuie să facă un creier este să selecteze semnale electrice care au doar legătură indirectă cu orice se întâmplă acolo, afară. Stimulii (inputs) senzoriali nu au etichete („eu provin

Scanned with OKEN Scanner

dintr-o ceașcă de cafea", „eu provin de la un arbore"). Nu au nici măcar etichete care să le sintetizeze natura – vizuală, auditivă, atingere sau unele mai puțin familiare, cum sunt termorecepția (senzația de temperatură) sau propriocepția (simțul care ne permite să ne percepem poziția corpului).*

Cum anume transformă creierul aceste semnale senzoriale, ambigui prin definiție, într-o lume coerentă perceptiv, plină de obiecte, oameni și locuri? În partea a doua a acestei cărți explorăm ideea potrivit căreia creierul este o „mașină de predicție", iar ceea ce vedem, auzim și simțim nu este altceva decât „cea-mai-bună-predicție" a creierului cu privire la cauzele input-urilor senzoriale. Urmând această idee până la capăt, vom constata despre conținuturile conștienței că sunt un fel de vis cu ochii deschiși – o halucinație controlată – care este mai mult, dar și mai puțin decât orice ar fi lumea reală.

Iată o viziune de bun simț asupra percepției. Să o numim varianta „cum par lucrurile".

Există, acolo afară, o realitate independentă de minte, plină de obiecte, oameni și locuri care chiar au proprietăți precum culoarea, forma, textura și așa mai departe. Simțurile noastre sunt ca niște ferestre transparente către această lume, detectând obiectele și trăsăturile lor și transmițând aceste informații către creier, acolo unde procesele neuronale complexe le descifrează, pentru a forma percepții. O ceașcă de cafea din acea lume duce la percepția unei căni de cafea, percepție generată în creier. Cât privește cine sau ce anume generează percepția – ei bine, acesta este „sinele", nu-i așa, acel „eu din spatele ochilor" („I behind the eye"), am putea spune, destinatarul valurilor succesive de date senzoriale,

* Ideea foarte cunoscută, dar complet greșită, potrivit căreia oamenii au doar cinci simțuri ,a fost formulată de Aristotel, în *De Anima* – „Despre suflet" – în jurul anului 350 î.Hr.

Scanned with OKEN Scanner

care își folosește interpretările perceptive pentru a-și ghida comportamentul, pentru a decide ce să facă în continuare. E o ceașcă de cafea acolo. O percep și o ridic. Simt, gândesc, apoi acționez.

Aceasta este o imagine atrăgătoare. Tiparele de gândire stabilite de-a lungul deceniilor, poate al secolelor, ne-au obișnuit cu ideea că creierul este un fel de computer acolo sus, în interiorul craniului, care procesează informații senzoriale pentru a construi o imagine interioară a lumii exterioare în folosul sinelui. Această imagine este atât de familiară, încât orice alternativă rezonabilă pare greu de conceput. Într-adevăr, mulți neurologi și psihologi încă gândesc percepția în acest fel, ca pe un proces de detectare a caracteristicilor „de jos în sus"/„bottom-up".

Iată cum se presupune că funcționează modelul bottom-up. Stimulii externi – unde luminoase, unde sonore, molecule care transmit gusturi și mirosuri șamd – declanșează impulsuri electrice în organele senzoriale, impulsuri care se transmit „în sus"/„up" sau „către înăuntru" /„inwards", spre creier. Aceste semnale senzoriale trec prin mai multe etape diferite de procesare, indicate de săgețile negre din imaginea de mai jos, fiecare etapă analizând caracteristici din ce în ce mai complexe. Să luăm ca exemplu vederea. Inițial, percepem caracteristici precum luminanța sau marginile unei persoane sau ale unui obiect,

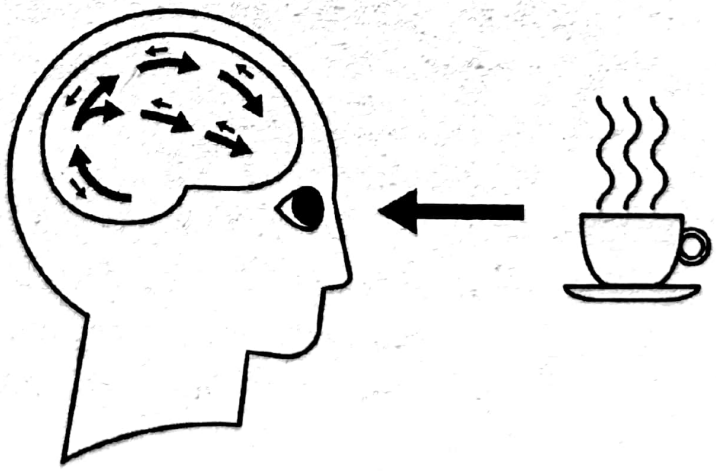

Fig. 3: Percepția ca detecție a caracteristicilor pe modelul bottom-up.

Scanned with OKEN Scanner

ulterior, aprofundând, trecem la părțile persoanei sau ale obiectului – cum ar fi ochii și urechile sau roțile și oglinzile laterale. Avansând în percepție, putem vedea persoanele/obiectele întregi sau categorii de obiecte, cum ar fi fețele, mașinile și ceștile de cafea.

În acest fel, lumea exterioară, cu obiectele și oamenii și tot-felul-de-altele este recapitulată sub forma unei serii de caracteristici extrase din râul de date senzoriale care se varsă în creier – caracteristici comparabile cu niște pești, din ce în ce mai mari și mai complecși, pe măsură ce râul se apropie de mare. Semnalele care curg în direcția opusă – de „sus în jos" („top down") sau „din interior în afară" („inside out"), micile săgeți gri – servesc doar pentru a rafina sau a constrânge într-un anume mod fluxul cel mai important, de jos în sus, al informațiilor senzoriale.

Această viziune bottom-up a percepției se potrivește perfect cu ceea ce știm despre anatomia creierului, cel puțin la prima vedere. În cortex, diferitele tipare perceptive sunt asociate cu regiuni specifice: cortexul vizual, cortexul auditiv șamd. În cadrul fiecărei regiuni, procesarea perceptivă este organizată ierarhic. În sistemul vizual, nivelurile inferioare, cum ar fi cortexul vizual primar, sunt aproape de input-urile senzoriale, în timp ce nivelurile superioare, cum ar fi cortexul temporal inferior, sunt cu câteva etape de procesare mai departe. În ceea ce privește conectivitatea, semnalele de la fiecare nivel sunt colectate la nivelul următor, astfel încât neuronii de la nivelurile superioare să poată răspunde la caracteristicile care pot fi distribuite în spațiu sau timp – așa cum ne-am aștepta.

Studiile asupra activității creierului par, de asemenea, să confirme varianta bottom-up. Experimentele de acum câteva decenii – care au investigat sistemele vizuale la pisici și maimuțe – au demonstrat în mod repetat că neuronii din primele stadii (inferioare) ale procesării vizuale răspund la caracteristici simple, cum ar fi marginile, în timp ce

Scanned with OKEN Scanner

neuronii din etapele ulterioare (superioare) răspund la trăsături complexe precum fețele. Experimente mai recente, care utilizează metode de neuroimagistică precum RMN funcțional, confirmă aproximativ același parcurs al evenimentelor în creierul uman.

Se pot construi chiar „sisteme de percepție" artificiale în acest fel – unele rudimentare, cel puțin. Teoria computațională clasică a vederii, formulată în 1982 de informaticianul David Marr, este atât un reper standard în abordarea bottom-up a percepției, cât și o rețetă practică pentru proiectarea și construcția sistemelor artificiale de vizualizare. Sistemele mai recente de vedere automată (machine vision), realizate folosind rețele neuronale artificiale – cum sunt rețelele de învățare profundă („deep learning") – ating, în prezent, niveluri de performanță impresionante, comparabile, în unele situații, cu ceea ce pot face oamenii. Și aceste sisteme se bazează adesea pe teoriile bottom-up.

În concluzie, percepția de tip bottom-up asupra a „cum par lucrurile" pare să fie deslușită destul de temeinic.

Ludwig Wittgenstein: „De ce se spune că e mai firesc să credem că Soarele se învârte în jurul Pământului, decât că Pământul se învârte în jurul axei lui?"

Elizabeth Anscombe: „Deoarece părea că soarele se învârte în jurul Pământului, presupun."

Ludwig Wittgenstein: „Atunci, cum ar fi fost dacă ar fi *părut că* Pământul se învârte în jurul axei lui?"

În acest dialog încântător dintre Wittgenstein și colega sa, Elizabeth Anscombe, filosof (și biograf), legendarul gânditor german folosește revoluția copernicană pentru a

Scanned with OKEN Scanner

ilustra faptul că felul în care *par* lucrurile nu este neapărat și felul în care chiar *sunt*. Deși *pare că* Soarele se învârte în jurul Pământului, Pământul este cel care, desigur, se rotește în jurul propriei axe, generând succesiunea a ceea ce noi numim noapte și zi, iar Soarele, nu Pământul, este situat în centrul sistemului solar. Nimic nou, ai putea crede, și ai avea dreptate. Dar Wittgenstein viza ceva mai profund. Ceea ce încerca să-i comunice lui Anscombe era că, chiar și înțelegând cum stau lucrurile, de la un anumit nivel, toate par să fie așa cum au fost mereu. Soarele răsare în est și apune în vest, la fel ca întotdeauna.

Cu percepția se întâmplă la fel ca în cazul sistemului solar. Deschid ochii și *pare că* există o lume reală acolo. Astăzi sunt acasă, în Brighton. Nu sunt chiparoși, ca în Santa Cruz, doar obiectele de fiecare zi împrăștiate pe birou, un scaun roșu în colț și, dincolo de geam, tot felul de coșuri de fum. Aceste obiecte *par să aibă* anumite forme și culori, iar cele mai apropiate, inclusiv mirosuri și texturi. Așa *par* lucrurile.

Deși poate *părea că* simțurile mele oferă ferestre transparente către o realitate independentă de minte, iar acea percepție este un proces de „descifrare" a datelor senzoriale, ceea ce se întâmplă cu adevărat este – cred eu – destul de diferit. Percepțiile nu vin de jos în sus (bottom-up) sau din exterior spre interior (outside-in), ci de sus în jos (top-down) sau din interior spre exterior (inside-out), în primul rând. Ceea ce trăim este construit din predicțiile creierului, „cele-mai-bune-presupuneri" despre cauzele semnalelor senzoriale. Ca și în cazul revoluției copernicane, viziunea top-down despre percepție rămâne în concordanță cu multe dintre dovezile existente, lăsând neschimbate multe aspecte ale modului în care par lucrurile, schimbând, însă, în același timp, totul.

Ideea nu e deloc nouă. Primele semne ale unei teorii top-down a percepției apar în Grecia antică, în Alegoria

Scanned with OKEN Scanner

Peșterii, a lui Platon, unde niște prizonieri înlănțuiți, toată viața lor, cu fața spre un perete gol, văd doar jocul umbrelor aruncate de obiectele care trec pe lângă un foc aprins în spatele lor. Dau nume umbrelor pentru că, pentru ei, ele sunt realitatea. Alegoria se referă la propriile noastre percepții conștiente, care sunt la fel ca acele umbre, reflexii indirecte ale unor cauze ascunse pe care nu le putem întâlni niciodată direct.

Mai mult de o mie de ani mai târziu, dar totuși cu o mie de ani în urmă, savantul arab Ibn al Haytham a scris că percepția, aici și acum, depinde mai degrabă de procesele de „judecată și inferență", decât de accesul direct la o realitate obiectivă. După încă câteva sute de ani, Immanuel Kant a realizat că datele senzoriale nerestricționate generează un haos care va rămâne întotdeauna lipsit de sens dacă nu e structurat de concepții preexistente care, pentru el, se referă la cadre *a priori*, precum spațiul și timpul. Termenul *noumen* al lui Kant se referă la „lucrurile în sine" – *Ding an sich* – o realitate independentă de minte, mereu inaccesibilă percepției umane, ascunsă în spatele unui văl senzorial.

În neuroștiință, povestea continuă cu fizicianul și fiziologul german Hermann von Helmholtz. La sfârșitul secolului al XIX-lea, Helmholtz a propus, printre alte contribuții remarcabile, ideea percepției ca proces de „inferență inconștientă". Conținutul percepției, susținea el, nu este dat de semnalele senzoriale în sine, ci trebuie *dedus* prin combinarea lor cu așteptările și/sau convingerile creierului despre cauzele lor. Considerând acest proces ca fiind inconștient, Helmholtz afirmă că noi nu suntem conștienți de mecanismele prin care au loc inferențele perceptive, ci doar de rezultate. Judecățile perceptive – „inferențele inconștiente" – țin evidența cauzelor respectivelor percepții, actualizând continuu și activ cele-mai-bune-presupuneri perceptive pe măsură ce se adună noi

Scanned with OKEN Scanner

date senzoriale. Helmholtz a considerat că oferă o versiune științifică a abordării lui Kant cum că percepția nu ne poate permite să cunoaștem lucrurile în mod direct - că putem doar deduce despre lucruri că sunt acolo, în spatele vălului senzorial.

Ideea fundamentală a lui Helmholtz despre „percepție ca inferență" a fost extrem de influentă, luând diverse forme pe parcursul secolului al XX-lea. În anii 1950, mișcarea „new look" din psihologie a subliniat influența factorilor sociali și culturali asupra percepției. De exemplu, conform unuia dintre cele mai citite studii la acea vreme, copiii din familii sărace supraestimau dimensiunea monedelor, comparativ cu copiii din familiile înstărite. Din păcate, multe experimente de acest fel – deși fascinante – erau făcute după metodologii mult sub standardele celor de azi, așa că rezultatele nu prea pot fi luate în considerare.

În anii '70, psihologul Richard Gregory a brodat în diverse feluri pe ideile lui Helmholtz, sugerând o teorie a percepției ca un fel de „testare de ipoteză" neuronală. Conform acestei teorii, așa cum cercetătorii testează și actualizează ipotezele științifice cu datele obținute din experimente, tot așa și creierul formulează continuu ipoteze perceptive despre cum este lumea – pe baza experiențelor trecute și a altor forme de informații stocate – și testează aceste ipoteze prin colectarea de date de la organele senzoriale. Conținutul perceptiv, pentru Gregory, este determinat de ipotezele cel mai bine susținute ale creierului.

După ce s-a tot estompat și reapărut pe parcursul ultimilor 50 de ani, ideea percepției ca inferență a reapărut cu forțe noi în ultimul deceniu. Au înflorit tot felul de teorii cu titluri generale de „codificare predictivă" și „procesare predictivă". Deși diferă în detalii, aceste teorii împărtășesc ideea comună conform căreia percepția depinde de un fel de inferență care are loc în creier.

Scanned with OKEN Scanner

Propria mea interpretare a ideii lui Helmholtz și a încarnărilor ei contemporane este cel mai bine surprinsă de noțiunea de percepție ca *halucinație controlată*, expresie pe care am auzit-o pentru prima dată de la psihologul britanic Chris Frith cu mulți ani în urmă.* Ingredientele esențiale ale halucinației controlate, așa cum o concep eu, sunt după cum urmează.

În primul rând, creierul prezice în mod constant cauzele semnalelor senzoriale, iar predicțiile sale se descarcă în cascadă de sus în jos/top-down prin ierarhiile perceptive cerebrale (săgețile gri din imaginea de mai jos). Atunci când privești o ceașcă de cafea, cortexul tău vizual va prezice cauzele semnalelor senzoriale care provin de la această ceașcă de cafea.

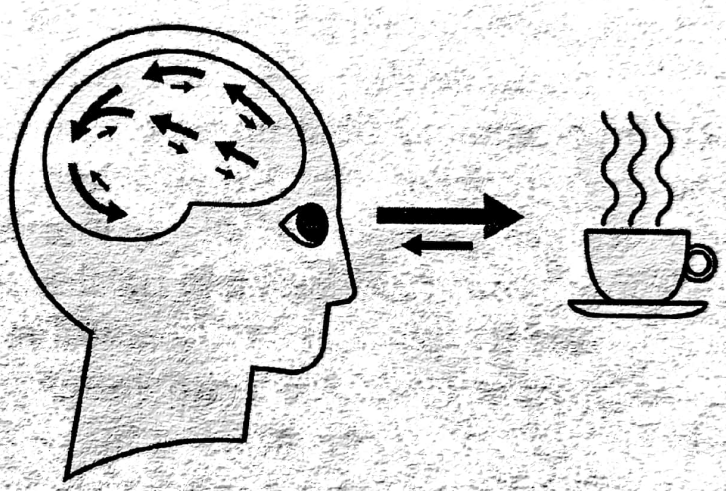

Fig. 4: Percepția ca inferență top-down.

În al doilea rând, semnalele senzoriale – colectate într-un flux de jos în sus/bottom-up sau din exterior spre interior/outside-in către creier – leagă aceste predicții perceptive, într-o manieră utilă, de cauzele lor: în acest caz, o ceașcă de cafea. Aceste semnale servesc ca *erori de predicție*, înregistrând diferența dintre ceea ce așteaptă creierul și ceea ce primește la fiecare nivel de procesare. Creierul reușește,

* Această expresie poate fi regăsită într-un seminar susținut în anii 1990 de Ramesh Jain. Am încercat să găsesc formularea și dincolo de acest an, dar fără succes.

Scanned with OKEN Scanner

astfel, să potrivească cele mai bune predicții ale sale cu cauzele semnalelor, prin ajustarea predicțiilor top-down, după suprimarea erorilor de predicție bottom-up. Din această perspectivă, percepția are loc printr-un proces continuu de *minimizare a erorilor de predicție.*

Al treilea și cel mai important ingredient în viziunea halucinației controlate este afirmația potrivit căreia experiența perceptivă – în acest caz, experiența subiectivă de „a vedea o ceașcă de cafea" – este determinată de conținutul predicțiilor (top-down), și nu de semnalele senzoriale (bottom-up). Nu experimentăm niciodată semnalele senzoriale în sine, experimentăm doar interpretări ale acestora.

Amestecând aceste ingrediente, pregătim, astfel, o inversare copernicană despre cum anume gândim percepția. *Pare că* lumea se dezvăluie direct minții noastre conștiente prin organele noastre de simț. Gândind astfel, este firesc să considerăm percepția ca fiind un proces bottom-up de detectare a caracteristicilor - o „citire" a lumii din jurul nostru. Dar ceea ce percepem, de fapt, este o fantezie neuronală de top-down și inside-out, în frâiele realității, nu o fereastră transparentă spre orice ar fi acea realitate.

Și – ca să-l parafrazez iar pe Wittgenstein – cum ar părea dacă *ar părea că* percepția ar fi cea mai bună presupunere top-down? Ei bine, așa cum soarele încă răsare la est și apune la vest, dacă *ar părea că* percepția ar fi o halucinație controlată, ceașca de cafea de pe masă – întreaga experiență perceptivă a oricui – ar părea la fel cum a părut și va părea întotdeauna.

Atunci când ne referim la halucinații, de obicei ne gândim la un fel de percepție generată intern, la vizualizarea sau la auzul a ceva ce nu e, de fapt, acolo – așa cum se poate întâmpla în schizofrenie sau în aventuri psihedelice precum cele ale lui Albert Hofmann. Aceste asocieri plasează halucinația în contrast cu percepția „normală", care se presupune că reflectă lucruri care există cu adevărat.

Scanned with OKEN Scanner

În perspectiva top-down a percepției, această distincție clară devine o chestiune de interpretare. Atât percepția „normală", cât și halucinația „anormală" implică predicții generate intern despre cauzele input-urilor senzoriale, ambele având în comun un set de bază de mecanisme cerebrale. Diferența este că, în percepția „normală", ceea ce percepem este legat de – *controlat de* – cauze din afară, în timp ce, în cazul halucinațiilor, percepțiile își pierd, într-o oarecare măsură, conexiunea cu aceste cauze. Când halucinăm, predicțiile noastre perceptive nu sunt actualizate corespunzător, în conformitate cu erorile de predicție.

Dacă percepția este o halucinație controlată, atunci – la fel – halucinația poate fi considerată o percepție necontrolată. Sunt diferite, dar a încerca să le delimitezi e ca și cum ai vrea să trasezi granița dintre zi și noapte.

Să analizăm un pic teoria halucinației controlate, pentru a afla ce înseamnă să experimentezi perceptiv *culoarea*.

Oricât de uimitor ar fi el, sistemul nostru vizual răspunde doar la o porțiune minusculă din întregul spectru electromagnetic, situată între infraroșul cu lungime mare de undă și ultravioletul cu lungime de undă mai mică. Fiecare culoare pe care o percepem, ba chiar fiecare parte a întregului fiecărei lumi personale vizuale, se bazează pe această felie subțire de realitate. Doar știind acest lucru este suficient să pricepem că experiența perceptivă nu poate fi o reprezentare cuprinzătoare a unei lumi obiective externe. Este mai puțin, dar și mai mult decât atât.

Întreabă un neurofiziolog – îți va spune că percepi o anumită culoare atunci când celulele retiniene cu con, sensibile la culoare, sunt activate într-o anumită proporție. Departe de a fi eronat, este, însă, departe de întreaga poveste. Nu există o corelare unu-la-unu între activarea

Scanned with OKEN Scanner

celulelor sensibile la culoare și experiența culorii. Culoarea pe care o vezi depinde de interacțiunea permanentă și complexă între lumina reflectată de suprafețe și iluminarea generală a mediului în care se întâmplă să te afli. Mai precis, depinde de modul în care creierul tău face inferențe – cele-mai-bune-presupuneri – despre modul în care se desfășoară această interacțiune.

Privește o bucată de hârtie albă, în lumina de afară. E tot albă, chiar dacă lumina pe care o reflectă acum are o compoziție spectrală foarte diferită, ca urmare a diferențelor dintre lumina (oarecum albăstruie) a soarelui și lumina (oarecum gălbuie) de interior. Sistemul tău vizual compensează automat aceste diferențe de iluminare ambientală – „reduce iluminantul", cum le place să spună cercetătorilor care studiază vederea – astfel încât culoarea pe care o vezi se bazează pe o proprietate invariabilă a hârtiei: *modul în care* reflectă lumina. Creierul deduce această proprietate invariantă și o consideră presupunerea sa cea mai bună referitoare la cauza input-urilor senzoriale în continuă schimbare. Gradul de alb este aspectul fenomenologic al acestei inferențe – este modul în care apar, în experiența noastră conștientă, deducțiile creierului despre această proprietate invariabilă.

Asta înseamnă, despre culoare, că nu este o proprietate definită a lucrurilor-în-sine. Culoarea este, mai degrabă, un dispozitiv util pe care evoluția l-a selectat astfel încât creierul să poată recunoaște și să țină evidența obiectelor în condiții de iluminare în schimbare. Trăirea experienței subiective a vederii scaunului roșu în colțul camerei nu înseamnă că scaunul *este de fapt roșu* – ce ar putea însemna, până la urmă, ca un scaun să aibă o proprietate fenomenologică precum roșul? Scaunele nu sunt roșii, așa cum nu sunt nici urâte, sau de modă veche, sau avangardiste. În schimb, suprafața scaunului are o proprietate anume, care e *modul-în-care-reflectă-lumina*, pe care creierul meu

Scanned with OKEN Scanner

o înregistrează prin mecanismele sale de percepție. Roșul este aspectul subiectiv, fenomenologic al acestui proces.

Înseamnă asta că roșul scaunului a trecut de la a fi „acolo", undeva, la „aici", în interiorul creierului? Într-un sens, răspunsul este, cu siguranță, „nu". Nu există roșu în creier în sensul naiv al existenței unui fel de pigment roșu – sau „iluzie/*figment* (în engleză, foarte aproape de *pigment*, n. trad.)" – în interiorul capului, care să fie inspectat de o cameră video în miniatură ce trimite un output către alt sistem vizual, la rându-i dotat cu o mini cameră video în interior... șamd. A presupune că o proprietate percepută a lumii exterioare (roșul) trebuie, cumva, să fie re-instanțiată în creier, pentru ca percepția să se întâmple, înseamnă să cazi victimă erorii numite a „dublei transducții", concept descris de filosoful Daniel Dennett. Conform acestei erori, un „roșu" extern este transdus de retină în modele de activitate electrică reconstituite apoi – transduse din nou – într-un „roșu" intern. Dennett susține că acest tip de raționament nu explică nimic. Singurul mod în care poate fi localizat roșul „în creier" este pur și simplu prin mecanismele care se găsesc la acel nivel și care stau la baza experienței perceptive. Aceste mecanisme, desigur, nu sunt roșii.

Când mă uit la un scaun roșu, roșul pe care îl trăiesc depinde atât de proprietățile scaunului, cât și de proprietățile creierului meu. Corespunde conținutului unui set de predicții perceptive despre modurile în care un anumit tip de suprafață reflectă lumina. Nu există roșu-ca-atare în lume sau în creier. Așa cum spunea Paul Cézanne, „culoarea este locul în care creierul nostru și universul se întâlnesc".

Avansând pe acest teritoriu, afirmația se aplică cu mult dincolo de tărâmul trăirii culorilor. Se aplică tuturor percepțiilor. Panorama imersivă multisenzorială a peisajului tău perceptiv, chiar aici și chiar acum, este una situată între creier și lume, o scriere, dar și o citire. Întregul experienței perceptive este o fantezie neuronală care rămâne cuplată

Scanned with OKEN Scanner

la exterior prin estimarea și re-estimarea celei mai bune predicții perceptive, a halucinațiilor controlate.

Ai putea spune chiar că toți halucinăm tot timpul. Abia atunci când vom fi de acord cu privire la halucinațiile noastre, vom trăi ceea ce numim realitate.

Iată trei exemple despre modul în care așteptările perceptive modelează experiența conștientă, exemple pe care le poți experimenta și singur.

Într-una dintre săptămânile lunii februarie din 2015, rețelele de socializare, dar și ziarele discutau intens despre „Rochie". În dimineața de miercuri a acelei săptămâni, ajuns în birou, m-a copleșit un potop de e-mailuri și mesaje vocale. Eram coautorul unei cărți despre iluziile vizuale, publicată recent, iar mass-media se chinuia să găsească explicații pentru un fenomen care părea că luase internetul cu asalt. Este vorba despre cunoscuta fotografie a unei rochii, colorate în albastru și negru, pentru unii, dar în alb și auriu, pentru alții.* Cei care au văzut-o într-una dintre cele două variante erau atât de convinși că au dreptate, încât nu le venea să creadă că cineva ar putea să o vadă altfel – și, astfel, internetul a explodat cu argumente și contraargumente.

La început, am crezut că e o farsă. Pentru mine, Rochia era evident albastră cu negru, la fel și pentru primii patru oameni din laborator cărora le-am arătat-o. A urmat surpriza, dar și ușurarea, când a cincea persoană a spus că rochia e albă cu auriu. În final, toți cei din laborator ne-am împărțit, la fel ca lumea întreagă, în două grupuri aproximativ egale – unul vedea rochia albastră cu negru, celălalt, albă cu auriu.

* Poți găsi o imagine color a Rochiei aici: https://en.wikipedia.org/wiki/The_dress. Ce vezi?

Scanned with OKEN Scanner

O oră mai târziu eram la BBC, încercând să explic ce se întâmpla. Apăruse un fel de consens general, cum că efectul asupra privitorului e legat de cantitatea de lumină care cade pe obiect – știut fiind că, la percepția culorilor, se ia în considerare lumina ambientală. Ideea era că lucrurile ar putea sta diferit pentru diferiți oameni, într-un mod mai puțin evident și necunoscut până atunci, dar care se întâmpla să facă diferența pentru Rochie.

Unii au susținut că fotografia Rochiei este supraexpusă și lipsită de context – rochia ocupă, singură, cea mai mare parte a imaginii – ceea ce ar putea păcăli creierul în procesul de generare a culorii din context. Dacă, dintr-un motiv oarecare, sistemul vizual este obișnuit cu lumina ambientală gălbuie – ca atunci când petreci prea mult timp în interior – atunci, ceea ce inferă este, mai probabil, albastru-negru. Dacă, dimpotrivă, ești o persoană de exterior, sănătoasă și fericită, cu un cortex vizual scăldat frecvent în lumina oarecum albastră a soarelui, atunci ai vedea alb-auriu.

Imediat după aceea, oamenii au început să facă tot felul de experimente: să se uite la fotografie într-o cameră slab luminată, repezindu-se apoi afară, la lumina zilei; corelând prevalența raportării alb-auriului cu ratele medii de însorire în diferite țări; verificând dacă bătrânii sunt mai predispuși să vadă albastru-negru decât tinerii. În scurt timp, a luat avânt o adevărată industrie de apartament pentru testări ca cele de mai sus sau a altor nenumărate ipoteze.

Faptul că oamenii treceau prin experiențe atât de diferite privind aceeași fotografie – și le și mărturiseau cu atâta încredere – reprezintă o dovadă convingătoare că experiențele noastre perceptive sunt constructe interne, modelate de idiosincraziile biologiei și istoriei noastre personale. În general, presupunem că vedem cu toții lumea aproximativ în același mod și, de cele mai multe ori, poate chiar așa e. Chiar dacă e așa, nu e pentru că scaunele roșii *sunt cu adevărat roșii*, ci pentru că creierul fiecăruia

Scanned with OKEN Scanner

dintre noi se bazează pe propriile estimări perceptive – iar o situație neobișnuită, cum a fost Rochia, ne ajută să înțelegem diferențele dintre noi.

Al doilea exemplu este îndrăgita iluzie vizuală numită Tabla de șah a lui Adelson. Acest exemplu arată că influența predicțiilor asupra percepției nu se limitează la situații ciudate, precum Rochia, ci se manifestă peste tot și tot timpul. Aruncă o privire la tabla de șah din partea stângă din imaginea de pe pagina alăturată și compară pătratele A și B. A pare mai întunecat decât B. Cel puțin, așa pare – mie și oricui le-am arătat vreodată. Niciun indiciu de diferențe individuale aici.

De fapt, A și B au exact aceeași nuanță de gri. Tabla de șah din dreapta demonstrează acest lucru prin unirea A și B cu un dreptunghi care are o nuanță uniformă de gri. Priviți cât de atent doriți, nu există modificări de umbrire, nici tranziții de orice fel. A și B au aceeași nuanță de gri, deși în tabla de șah din stânga continuă să arate diferit. Faptul că știi că sunt la fel nu ajută. M-am uitat la aceste imagini de mii de ori, iar A și B (în stânga) se încăpățânează să aibă nuanțe diferite de gri.*

Ceea ce se întâmplă aici este că percepția cenușiului este determinată nu de undele luminoase reale care vin de la A sau B – și care sunt aceleași – ci de cea mai bună presupunere a creierului despre ce anume a cauzat aceste combinații particulare de lungimi de undă și – ca și în cazul Rochiei – asta depinde de context. B este în umbră, A nu este, iar sistemul vizual al creierului are bine întipărită

* Atunci când cunoașterea nu reușește să afecteze percepția, numim percepția „impenetrabilă din punct de vedere cognitiv".

Scanned with OKEN Scanner

în circuitele sale ideea că obiectele din umbră par mai întunecate. Creierul își ajustează deducțiile despre umbra lui B pe baza cunoașterii anterioare despre umbre, la fel cum își ajustează deducțiile perceptive pe baza luminii ambientale. Acesta este motivul pentru care, pe tabla de șah din partea stângă, percepem B ca fiind mai deschis decât A (unde nu este umbră). În schimb, pentru tabla de șah din dreapta, contextul de umbră este perturbat de cele două bare gri așezate pe deasupra, astfel încât putem vedea că A și B sunt, de fapt, identice.

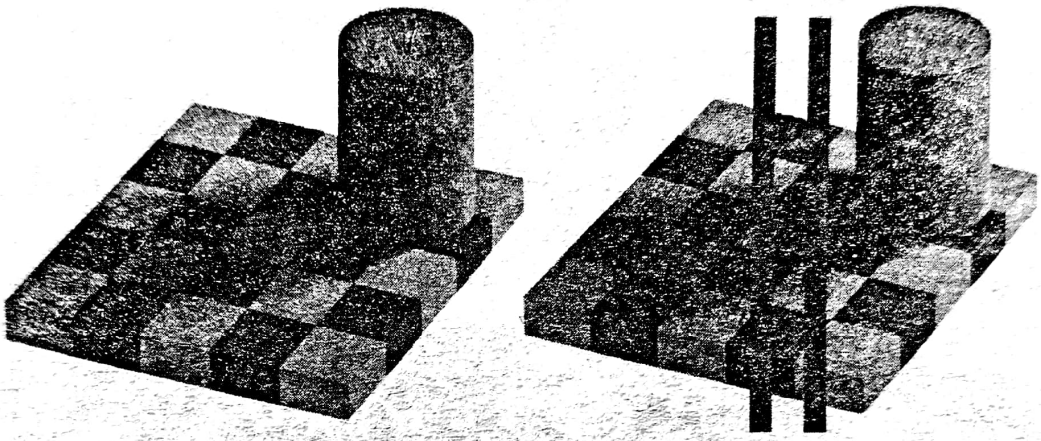

Fig. 5: Tabla de șah a lui Adelson.

Toate acestea sunt complet automate. Nu ești – sau cel puțin nu erai – conștient de faptul că creierul tău are, dar și folosește așteptările precedente despre umbre atunci când își face predicțiile perceptive. Nu este nici defect al sistemului vizual. Nu se pune problema ca un sistem vizual util să fie un exponometru, de tipul celor folosite de fotografi. Funcția percepției, cel puțin la o primă aproximare, este de a identifica cele mai probabile cauze ale semnalelor senzoriale, nu de a furniza însăși conștientizarea semnalelor senzoriale – orice ar însemna asta.

Scanned with OKEN Scanner

Exemplul final dezvăluie cât de repede noile predicții pot influența percepția conștientă. Privește imaginea de mai jos. Probabil tot ce poți vedea este o amestecătură de pete negre și albe. Apoi, după ce ai citit și restul acestei propoziții, privește imaginea de la pag. 98 și revino aici.

Fig. 6: Ce este asta?

Bun, te-ai întors. Acum privește din nou imaginea de pe această pagină; ar trebui să arate destul de diferit. Acolo unde înainte era o un amestec de pete, acum există obiecte distincte, *lucruri* – ceva se întâmplă. Aceasta este o „imagine în două tonuri" sau „imagine Mooney". Odată văzută, este greu de uitat. Imaginile în două tonuri sunt create făcând o fotografie, redând-o în tonuri de gri și modificând-o apoi cu atenție, astfel încât detaliile să se piardă la extremitățile albului și negrului – „cele două tonuri". Făcută în mod corect și cu imaginea potrivită, devine foarte greu să-ți dai seama de original. Asta până vezi originalul, caz în care imaginea în două tonuri se transformă brusc într-o scenă coerentă.

Scanned with OKEN Scanner

Remarcabil la acest exemplu este că, dacă te uiți acum la imaginea originală în două tonuri, semnalele senzoriale care ajung la ochi nu s-au schimbat deloc față de prima dată când ai văzut-o. Tot ceea ce s-a schimbat sunt predicțiile creierului tău despre cauzele acestor date senzoriale și asta schimbă ceea ce vezi în mod conștient.

Acest fenomen nu caracterizează doar vederea. Există și exemple auditive convingătoare, cunoscute sub denumirea de „vorbire sinusoidală". În acest caz, o frază rostită este procesată prin eliminarea tuturor frecvențelor înalte, frecvențe datorită cărora vorbirea normală devine inteligibilă. Rezultatul sună, de obicei, ca un șuierat zgomotos, total lipsit de sens – echivalentul auditiv al unei imagini în două tonuri. Dacă asculți fraza originală neprocesată, apoi, din nou, versiunea „undă sinusoidală", dintr-o dată totul devine clar. La fel ca și în cazul imaginilor în două tonuri, o predicție clară despre cauzele semnalelor senzoriale modifică – îmbogățește – experiența perceptivă.

Împreună, aceste exemple – deși, ce-i drept, deliberat simple – arată că percepția este un act generativ, creativ; o interpretare proactivă, încărcată de context, dar și implicată cu semnalele senzoriale. Și, așa cum am menționat mai devreme, principiul conform căruia experiența perceptivă este construită pe baza predicțiilor creierului se aplică peste tot – nu numai vederii și auzului, ci tuturor percepțiilor noastre, tot timpul.

O implicație importantă a acestui principiu este aceea că nu experimentăm niciodată lumea „așa cum este". Așa cum a subliniat Kant cu noumenul său, e greu de știut ce ar însemna asta cu adevărat. Chiar și ceva atât de

Scanned with OKEN Scanner

elementar precum culoarea există, așa cum am văzut, doar în interacțiunea dintre o lume și o minte. Așadar, deși am putea fi surprinși când iluziile perceptive – precum cele abia descrise – dezvăluie o discrepanță între ceea ce vedem (ori auzim, ori atingem) și lucrurile exterioare, ar trebui să nu judecăm experiențele perceptive doar în funcție de „acuratețea" cu care coincid cu realitatea. Percepția exactă – „veridică" –, înțeleasă astfel, este o himeră. Halucinația controlată a lumii noastre perceptive a fost proiectată, prin evoluție, să ne sporească șansele de supraviețuire, nu pentru a fi o fereastră transparentă către o realitate externă, o fereastră care, oricum, nu are sens conceptual. Vom aprofunda aceste idei în capitolele următoare, dar, înainte de asta, merită să comentăm câteva obiecții.

Prima obiecție este că perspectiva halucinației controlate a percepției neagă aspecte incontestabile ale lumii reale. „Dacă tot ceea ce trăim este doar un fel de halucinație", ai putea zice, „atunci sari în fața unui tren și vezi ce se întâmplă."

Nimic din ceea ce spun nu ar trebui folosit pentru a nega existența lucrurilor în lume, fie ele trenuri în viteză, pisici sau cești de cafea. „Controlul", în halucinația controlată, este la fel de important ca și „halucinația". A descrie percepția în acest fel nu înseamnă că orice e acceptabil, ci că modul în care ne *apar* lucrurile în experiența perceptivă a lumii este un construct al creierului.

Acestea fiind spuse, este util să facem diferența dintre calitățile „primare" și cele „secundare" – numite astfel de filosoful iluminist John Locke. După Locke, calitățile primare ale unui obiect sunt cele care există independent de un observator, cum ar fi faptul că ocupă spațiu, sunt solide sau se mișcă. Un tren care se apropie are o grămadă de astfel de calități primare, motiv pentru care să sari în fața unuia este o idee proastă, indiferent dacă îl observi sau nu și indiferent de ceea ce crezi despre natura percepției.

Scanned with OKEN Scanner

Calitățile secundare sunt cele a căror existență depinde de un observator. Acestea sunt proprietăți ale obiectelor care produc senzații – sau „idei" – în minte și nu se poate spune că există independent în obiect. Culoarea este un bun exemplu de calitate secundară, deoarece experiența culorii depinde de interacțiunea unui anumit tip de aparat perceptiv cu un obiect.

Din perspectiva halucinației controlate, atât calitățile primare, cât și cele secundare ale obiectelor pot da naștere experienței perceptive printr-un proces activ, constructiv. Conținutul experienței perceptive nu este, însă, în niciunul dintre cazuri, identic cu calitatea corespunzătoare a obiectului.

A doua obiecție are de-a face cu capacitatea noastră de a percepe lucruri noi. Ne-am putea îngrijora că avem nevoie de o cea-mai-bună-presupunere pre-formată pentru orice am putea percepe vreodată, astfel că am rămâne captivi într-o lume perceptivă a deja-așteptatului. Imaginează-ți că nu ai văzut niciodată o gorilă, nici în viața reală, nici la TV sau într-un film – nici măcar într-o carte – și apoi te întâlnești cu una pe stradă. Îți garantez că în acel moment vezi gorila – experiență perceptivă nouă și, probabil, mai degrabă înfricoșătoare. Într-o lume a lucrurilor deja-așteptate, cum se poate întâmpla asta?

Iată răspunsul scurt: „a vedea o gorilă" nu este niciodată o experiență perceptivă complet nouă. Gorilele sunt animale cu brațe, picioare și blană, iar tu – și strămoșii tăi – ai mai văzut și alte creaturi cu unele sau cu toate aceste trăsături. Mai general, gorilele sunt ceva cu margini definite (deși blănoase), care se mișcă în moduri rezonabil previzibile și care reflectă lumina în același mod în care o fac alte lucruri de dimensiuni, culoare și textură similare. Noua experiență de a „vedea o gorilă" este construită din predicții perceptive care operează pe mai multe niveluri diferite de granularitate și dobândite pe mai multe scale

Scanned with OKEN Scanner

de timp diferite – de la predicții despre luminanță și margini, până la predicții despre fețe și postură – care, împreună, formează cea mai bună predicție perceptivă, astfel că ajungi să vezi o gorilă pentru prima dată.

Răspunsul mai lung cere să aflăm mai multe despre modul în care creierul efectuează gimnastica neuronală extrem de complexă implicată în inferența perceptivă – și către acest subiect ne îndreptăm în următorul capitol.

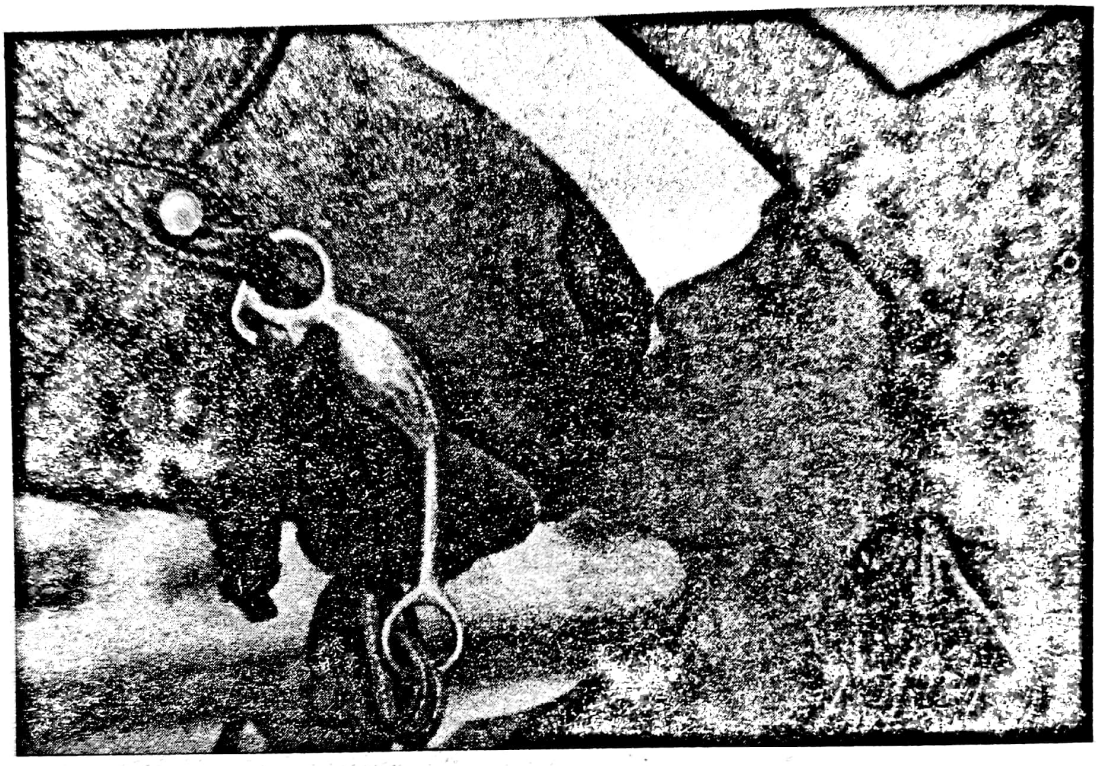

Fig. 7: Acesta este ce. Imagine de referință pentru fig. 6.*

* Figurile 6 și 7 din Teufel, C., Dakin, S. C. & Fletcher, P. C. (2018), „Prior object-knowledge sharpens properties of early visual feature detectors" (Cunoașterea anterioară a obiectelor accentuează proprietățile detectorilor de caracteristici vizuale timpurii), *Scientific Reports*, 8:10853. Folosit cu permisiunea autorilor și sub o licență internațională Creative Commons Attribution 4.0, cu mulțumiri lui Christoph Teufel.

Scanned with OKEN Scanner

5
Vrăjitorul din Odds*

Reverendul Thomas Bayes (1702 – 1761) – preot presbiterian, filosof și statistician, care și-a trăit cea mai mare parte a vieții în Tunbridge Wells, sudul Angliei – nu a reușit niciodată să-și publice teorema care i-a imortalizat numele. Al său „Essay towards Solving a Problem in the Doctrine of Chances" (Eseu despre rezolvarea unei probleme în doctrina șanselor) a fost prezentat la Royal Society din Londra, la doi ani după moartea sa, de către un coleg predicator, filosoful Richard Price, iar cea mai mare parte a demonstrației matematice aparține matematicianului francez Pierre-Simon Laplace. Dar numele lui Bayes este cel care va marca pentru totdeauna acest mod de gândire numit „inferență pentru cea mai bună explicație", ale cărui perspective sunt esențiale pentru înțelegerea modului în care percepția conștientă se construiește pe cele mai bune estimări ale creierului.

Raționamentul bayesian se referă în întregime la raționamentul cu probabilități. Mai precis, se referă la modul în care facem inferențe optime – ceea ce am numit până acum „cele mai bune estimări/predicții" – în condiții de incertitudine. „Inferență", un termen cu care ne-am întâlnit deja, înseamnă a ajunge la o concluzie pe bază de dovezi și de rațiune. Inferența bayesiană este un exemplu de raționament *abductiv*, care este diferit de cel *deductiv* sau de cel *inductiv*. A deduce înseamnă a ajunge la o concluzie doar pe baza logicii: dacă Jim este mai în vârstă decât Jane, iar Jane este mai în vârstă decât Joe, atunci Jim este mai

* Joc de cuvinte: Vrăjitorul din Oz devine, aici, din Odds – unde odds înseamnă probabilitate /șansă, în statistică.

Scanned with OKEN Scanner

în vârstă decât Joe. Dacă premisele sunt adevărate și se respectă regulile logicii, atunci inferențele deductive sunt, fără îndoială, corecte. Raționamentul inductiv se referă la maniera de a ajunge la o concluzie prin extrapolare de la o serie de observații: dacă soarele a răsărit la est pe tot parcursul istoriei consemnate, atunci a răsărit mereu de la est.

Spre deosebire de inferențele deductive, cele inductive pot fi eronate: dacă primele bile extrase dintr-un vas sunt verzi, atunci toate bilele din vas sunt verzi. Această afirmație poate fi adevărată sau nu.

Raționamentul abductiv – de tipul celui formalizat de inferența bayesiană – se referă la găsirea celei mai bune explicații pentru un set de observații, atunci când acele observații sunt incomplete, nesigure sau ambigui într-o manieră oarecare. Ca și raționamentul inductiv, cel abductiv poate fi eronat. În căutarea „celei mai bune explicații", raționamentul abductiv poate fi privit ca fiind mai degrabă unul în sens invers, de la efectele observate către cauzele lor cele mai evidente, decât de la cauză la efect – așa cum se întâmplă cu deducția și inducția.

Iată un exemplu. Privind pe fereastra dormitorului, într-o dimineață, vezi că pajiștea e udă. A plouat peste noapte? Poate, dar o altă variantă ar fi că ai uitat aspersoarele pornite. Ideea e să găsești cea mai bună explicație sau ipoteză pentru ceea ce vezi: *dat fiind* că iarba e udă, care este probabilitatea (i) să fi plouat peste noapte sau (ii) să fi uitat aspersoarele pornite? Cu alte cuvinte, vrem *să inferăm cauza cea mai probabilă a datelor observate*.

Inferența bayesiană ne spune cum să facem asta. Oferă calea optimă de actualizare a convingerilor noastre despre ceva, atunci când apar date noi. Regula lui Bayes este o rețetă matematică de a ajunge de la ceva ce știm deja („precedent"), la ceea ce ar trebui să știm după („posterior"), pe baza a ceea ce aflăm acum („probabilitate"). Precedentele, posterioarele și probabilitățile se numesc

Scanned with OKEN Scanner

adesea „convingeri" bayesiene deoarece reprezintă, mai degrabă, stări ale cunoașterii, decât stări ale lumii. (De remarcat că o convingere bayesiană nu este neapărat ceea ce cred eu-ca-persoană. Are la fel de mult sens și dacă spun despre cortexul meu vizual că „crede" că obiectul din fața mea e o ceașcă de cafea, și dacă spun că cred că Neil Armstrong a ajuns pe lună.)

Precedentele sunt probabilitățile ca ceva să se întâmple înainte de a avea noile date. Să spunem că probabilitatea precedentă a unei ploi de noapte este foarte mică – poate te afli în Las Vegas. Probabilitatea precedentă să fi uitat aspersoarele deschise depinde de frecvența cu care le folosești sau de cât de uituc ești. Este tot mică, dar nu la fel ca probabilitatea precedentă de a ploua.

Probabilitățile sunt, în linii mari, opusul posterioarelor. Formalizează raționamentul dinspre cauză spre efect: date fiind ploaia de azi-noapte sau funcționarea aspersoarelor, care e probabilitatea ca iarba să fie udă? La fel ca precedentele, pot varia, dar acum să presupunem că ploaia și aspersoarele uitate deschise sunt responsabile în mod egal de iarba udă.

Regula lui Bayes combină precedentele și probabilitățile pentru a ajunge la o probabilitate posterioară pentru fiecare ipoteză. Legea în sine este simplă: posteriorul este precedentul înmulțit cu probabilitatea, împărțit de un al doilea precedent (acesta este „precedentul despre date" – care, în acest caz, este probabilitatea precedentă ca iarba să fie udă – nu trebuie să ne facem griji despre asta, cât timp este aceeași pentru fiecare ipoteză).

Observând iarba udă dimineața, un bayesian ar trebui să aleagă ipoteza cu cea mai mare probabilitate precedentă – aceasta fiind explicația cea mai probabilă a datelor. Cât timp, în exemplul nostru, probabilitatea precedentă a ploii peste noapte este mai mică decât cea a aspersoarelor rămase funcționale, probabilitatea posterioară a ploii va fi, și ea,

Scanned with OKEN Scanner

mai mică. Un bayesian va alege, deci, ipoteza aspersoarelor uitate deschise. Această ipoteză este presupunerea bayesiană cea mai bună a cauzelor datelor observate – este „inferența celei mai bune explicații".

Cea mai mare parte din ce am relatat pare de bun simț deoarece, în acest caz, chiar este. Există, însă, numeroase situații în care inferența bayesiană se îndepărtează de ceea ce sugerează bunul simț. De exemplu, este ușor să ajungi la concluzia că suferi de o boală urâtă pornind de la rezultatele unor teste medicale, din cauza tendinței noastre de a supraestima probabilitatea precedentă de a avea o boală rară. Chiar dacă un test are o acuratețe de 99%, un rezultat pozitiv crește foarte puțin probabilitatea posterioară de a avea boala asociată atunci când prevalența, în populație, este suficient de scăzută.

Să ne întoarcem la povestea cu pajiștea udă și să avansăm un pic mai departe. După ce observi iarba udă, privești la pajiștea vecinului și vezi că și aceea e udă. Aceasta este o informație semnificativ nouă. Probabilitatea pentru fiecare ipoteză este, acum, diferită: în cazul ipotezei aspersoarelor funcționale, doar pajiștea ta trebuia să fie udă, dar în cazul ipotezei ploii, ambele pajiști trebuie că sunt ude. (Probabilitățile, să ne amintim, se derulează dinspre presupusele cauze către datele observate.) Fiind un bayesian, actualizezi probabilitățile posterioare și ajungi la concluzia că ploaia din timpul nopții este acum cea mai bună explicație pentru ceea ce vezi – așa că-ți schimbi părerea.

O caracteristică semnificativă a inferenței bayesiene este aceea că ia în calcul *fiabilitatea* informației atunci când actualizează cea mai bună predicție. Informația care este (estimată a fi) fiabilă ar trebui să aibă o influență mai mare asupra convingerilor bayesiene decât informația care nu este (estimată a fi) fiabilă. Dacă geamul dormitorului ar fi murdar și/sau dacă nu-ți găsești ochelarii, pare că pajiștea vecinului ar putea fi udă, numai că noua informație este

Scanned with OKEN Scanner

nefiabilă, iar tu ești conștient de asta. În acest caz, deși ipoteza ploii de peste noapte pare ceva mai probabilă atunci când privești pe fereastră, ipoteza originală a aspersoarelor uitate deschise ar putea să prevaleze.

În multe situații, procesul de actualizare a celor mai bune estimări bayesiene cu date noi se întâmplă iar și iar, într-un ciclu fără sfârșit al inferenței. La fiecare iterație, posteriorul anterior devine noul precedent. Acest nou precedent este, apoi, folosit pentru a interpreta următorul set de date și a genera un nou posterior – o nouă cea mai bună estimare – iar ciclul se repetă. Dacă pajiștea ta e udă două dimineți la rând, atunci cea mai bună previziune referitoare la cauza din a doua zi ar trebui să fie informată de cea mai bună previziune din prima zi șamd, pe măsură ce trec zilele.

Aplicarea inferenței bayesiene a adus beneficii remarcabile în toate tipurile de contexte, de la diagnosticul medical până la căutarea de submarine dispărute, noi aplicații apărând tot timpul. Metoda științifică în sine poate fi înțeleasă ca un proces bayesian, în care ipotezele științifice sunt actualizate cu noi dovezi experimentale. O astfel de abordare diferă atât de „schimbarea de paradigmă" a lui Thomas Kuhn, în care edificii științifice întregi sunt răsturnate prin acumularea de date nefiabile, cât și de vederile „falsificaționiste" ale lui Karl Popper, în care ipotezele sunt formulate și testate una câte una, asemenea unor baloane eliberate spre cer și, apoi, doborâte. În filosofia științei, perspectiva bayesiană are cele mai multe în comun cu vederile filosofului ungur Imre Lakatos, a cărui analiză se concentrează mai degrabă pe ce-anume-face-ca programele de cercetare științifică să se aplice și în practică, decât pe în-ce-anume-ar-putea-să-constea, la modul ideal.

O perspectivă bayesiană a științei înseamnă, firește, că acele convingeri precedente ale oamenilor de știință despre validitatea teoriilor lor vor influența proporțiile până la

Scanned with OKEN Scanner

care aceste teorii sunt actualizate sau demontate de noile date. De exemplu, sunt conștient de convingerea mea solidă conform căreia creierele sunt automate de predicție de tip bayesian. Această convingerea solidă va modela nu numai maniera în care interpretez dovezile experimentale, ci va determina și tipurile de experimente pe care le fac, pentru a genera noi dovezi relevante pentru convingerile mele. Uneori mă întreb câte dovezi ar fi necesare pentru a-mi răsturna convingerea bayesiană că creierul este fundamental bayesian.

Să revenim la creierul pe care ni l-am imaginat, tăcut, în întunericul din interiorul craniului, încercând să-și dea seama ce se află acolo, afară. Este o ocazie ideală pentru a invoca inferența bayesiană. Atunci când face predicțiile cele mai bune despre cauzele semnalelor senzoriale, atât de zgomotoase și ambigui, creierul respectă principiile reverendului Thomas Bayes.

Precedentele perceptive pot fi codificate la multe niveluri de abstractizare și flexibilitate. Acestea se întind de la precedente foarte generale și oarecum fixe, cum ar fi „lumina vine de deasupra", la precedente specifice anumitor situații, precum „obiectul păros care se apropie este o gorilă". În creier, probabilitățile codifică sub formă de hărți – de la potențiale cauze, la semnale senzoriale. Acestea sunt componentele „raționamentului în avans" ale inferenței perceptive și, la fel ca și cu precedentele, pot opera pe diferite scale de timp și spațiu. Creierul combină continuu aceste precedente cu probabilitățile conform regulii lui Bayes, astfel că, în fiecare fracțiune de secundă, se formează un nou bayesian posterior – o cea-mai-bună-estimare. Iar fiecare nou posterior servește ca precedent pentru următoarea rundă de input senzo-

Scanned with OKEN Scanner

rial permanent în schimbare. Percepția este un proces dinamic, nu static.

Fiabilitatea informației senzoriale are și ea un rol important aici. Cu excepția situației în care ești la zoo, probabilitatea precedentă de a întâlni o gorilă este foarte mică atunci când sesizezi ceva inform și păros în depărtare. Deoarece indiferent-de-ce-ar-fi e la distanță, fiabilitatea estimată a input-ului vizual va fi și ea mică, astfel că estimarea ta cea mai bună nu va fi cea de „gorilă". Pe măsură ce animalul se apropie, însă, semnalele vizuale devin tot mai fiabile și mai informative, astfel că cele mai bune estimări ale creierului tău se vor deplasa către o serie de opțiuni – câine negru mare, om în costum de gorilă, gorilă adevărată – până când devii convins că ceea ce percepi este chiar o gorilă, în timp suficient de scurt pentru a o rupe la fugă.

Cel mai simplu mod de a gândi despre convingerile bayesiene – precedente, probabilități și posterioare – este sub formă de valori discrete între 0 (reprezentând probabilitatea zero) și 1 (reprezentând probabilitate 100%). Totuși, pentru a înțelege cum anume fiabilitatea semnalelor senzoriale influențează inferența perceptivă, dar și a vedea cum anume legea lui Bayes poate fi implementată, de fapt, în cazul creierului, trebuie să aprofundăm și să gândim, în schimb, în termeni de *distribuție a probabilității*.

Diagrama de pe pagina următoare este un exemplu de distribuție a probabilității pentru variabila X. O variabilă, în matematică, este un simbol care poate lua diferite valori. Distribuția probabilității pentru X descrie probabilitatea ca valoarea lui X să se afle într-un anumit domeniu. Așa cum arată diagrama, poate fi reprezentată printr-o curbă. Probabilitatea ca X să se afle într-un anumit domeniu este dată de aria de sub curbă ce corespunde acelui domeniu. În acest exemplu, probabilitatea ca X să se afle între doi și patru este mult mai mare ca probabilitatea ca X să se afle între patru și șase. La fel ca în cazul tuturor distribuțiilor

Scanned with OKEN Scanner

de probabilități, aria totală de sub curbă este unu. Aceasta deoarece, atunci când sunt luate în calcul toate rezultatele posibile, ceva trebuie să se întâmple.

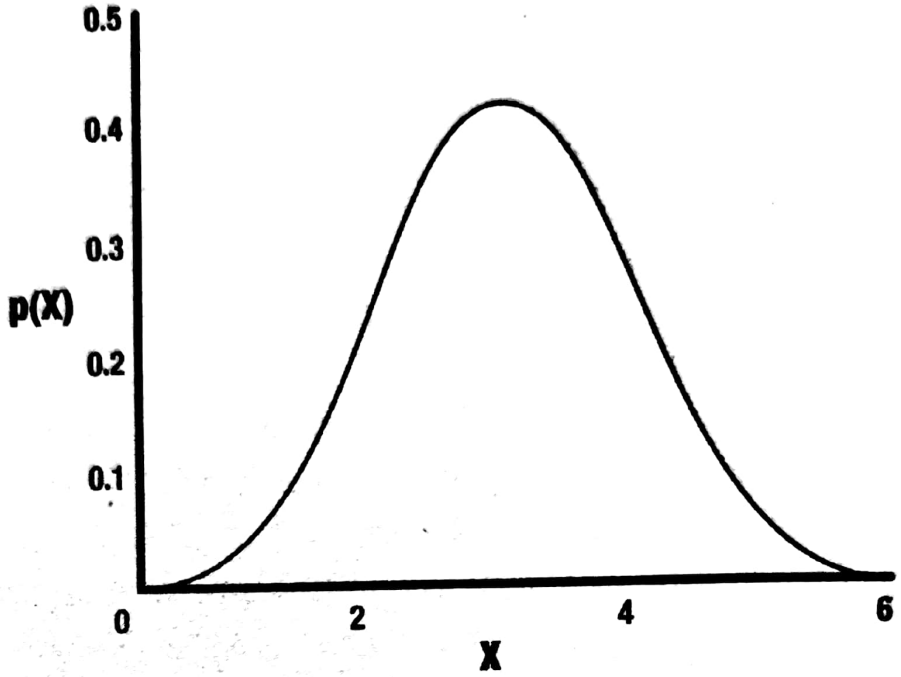

Fig. 8: Distribuția gaussiană a probabilității

Distribuția probabilității poate lua multe forme. O familie obișnuită de astfel de forme, din care face parte curba de mai sus, este distribuția „normală", „gaussiană" sau „în formă de clopot". Aceste distribuții sunt descrise complet de o valoare medie sau *medie* (acolo unde se află vârful curbei, în cazul de mai sus, 3) și de o *precizie* (cât de împrăștiată este; cu cât precizia este mai înaltă, cu atât împrăștierea e mai mică). Aceste cantități – media și precizia – se numesc *parametrii* distribuției*.

Convingerile bayesiene pot fi reprezentate, deci, prin distribuția gaussiană de acest tip. Intuitiv, media se referă la conținutul convingerii, iar precizia, la încrederea cu care creierul își menține convingerea. O distribuție ascuțită (precizie înaltă) corespunde unei încrederi foarte mari în respectiva convingere. Așa cum vom vedea, capacitatea de

* Uneori, în loc de precizie se folosește *varianță*. Precizia este inversul variației: cu cât precizia e mai mare, cu atât varianța este mai scăzută.

Scanned with OKEN Scanner

a reprezenta încrederea – sau fiabilitatea – este cea care dă greutate inferenței bayesiene.

Să revenim la exemplul cu gorila. Precedentele, probabilitățile și posterioarele relevante pot fi gândite acum ca distribuții de probabilitate, fiecare definită de o medie și de o precizie. Pentru fiecare distribuție, media semnifică probabilitatea „gorilei", iar precizia corespunde încrederii pe care creierul o are în estimarea probabilității.

Ce se întâmplă atunci când apar noi date senzoriale? E mai simplu să vizualizăm procesul de actualizare bayesiană. În următoarea diagramă, curba punctată reprezintă probabilitatea precedentă de a întâlni gorila. Această curbă are media mică, indicând faptul că gorilele se întâlnesc rareori, și o precizie destul de mare, indicând faptul că această convingere anterioară este susținută cu mare încredere. Curba întreruptă este probabilitatea, corespunzătoare input-ului senzorial. Aici, media e mai mare, dar precizia, mai mică: dacă chiar e vorba de o gorilă, atunci acestea sunt datele senzoriale pe care le-ai

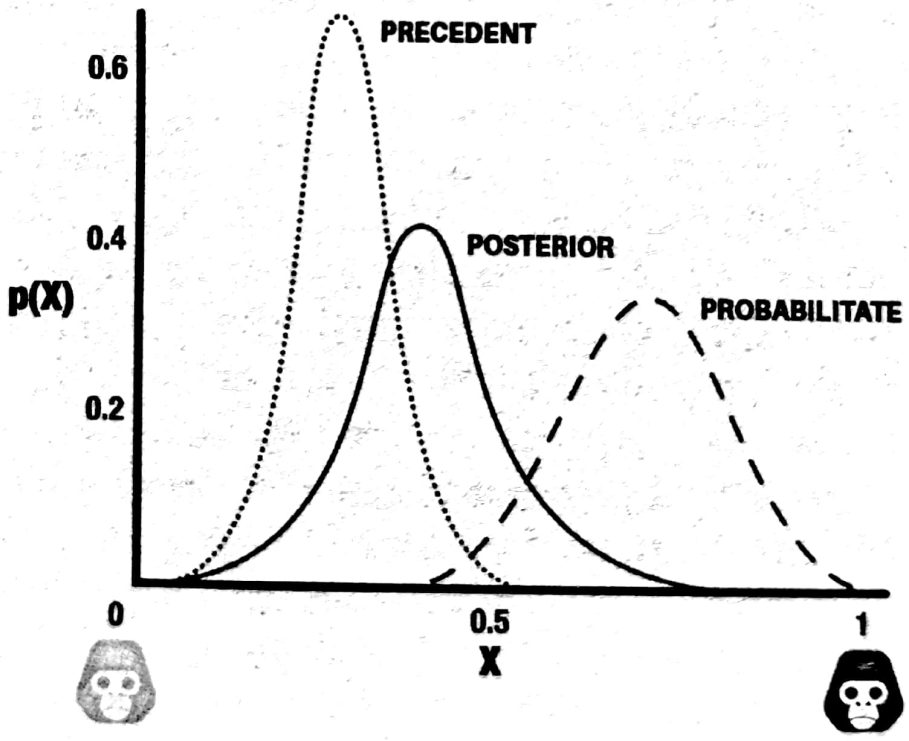

Fig. 9: Inferență bayesiană cu distribuție gaussiană a probabilităților pentru a estima cât mai bine prezența gorilei.

Scanned with OKEN Scanner

primit, numai că nu ești prea sigur că chiar așa e. Curba neîntreruptă este posterioara, reprezentând probabilitatea ca acolo să fie o gorilă, date fiind input-urile senzoriale.

Asta se obține, de obicei, prin aplicarea regulii lui Bayes. Atunci când avem de a face cu distribuția gaussiană de probabilități, aplicarea regulii lui Bayes înseamnă înmulțirea celor două curbe – punctată și întreruptă – păstrând aria de sub curba rezultată – posterioara – la limita 1.

De remarcat faptul că maxima posterioarei este mai aproape de precedentă decât de probabilitate. Aceasta deoarece combinarea celor două distribuții gaussiene depinde atât de medie, cât și de precizie. În acest caz, pentru că probabilitatea are precizie relativ scăzută – semnalele senzoriale ce indică „gorila" sunt estimate ca fiind nefiabile – cea mai bună estimare a posterioarei nu s-a deplasat foarte departe de precedentă. Totuși, la a doua privire, datele senzoriale de la gorilă par a fi ceva mai clare pentru că sunt, acum, mai aproape de tine, noua precedentă este dată de posterioara anterioară, astfel că noua posterioară – noua cea mai bună estimare – se deplasează mai aproape de „gorilă". Și tot așa, până când e timpul să o iei la goană.

Teorema lui Bayes oferă un standard de optim pentru inferența perceptivă. Prezintă scenariile optime pentru ceea ce ar trebui să facă creierul atunci când se străduiește să identifice cele mai probabile cauze ale input-urilor senzoriale, fie că acestea provin de la gorile, de la scaune roșii sau de la cești de cafea. Dar aceasta este doar o parte a poveștii. Ceea ce nu specifică teorema lui Bayes este tocmai *cum anume* reușește creierul să estimeze cât mai bine, în termeni de mecanisme neuronale.

Răspunsul la această problemă ne duce înapoi la teoria halucinației controlate despre percepție și la afirmația principală conform căreia conținuturile conștiente nu sunt modelate doar de predicții perceptive – conținuturile conștiente *sunt* aceste predicții.

Scanned with OKEN Scanner

În capitolul anterior am introdus ideea conform căreia percepția are loc printr-un proces continuu de *minimizare a erorii de predicție*. Astfel, creierul generează continuu predicții despre semnalele senzoriale și compară aceste predicții cu semnalele senzoriale care ajung la ochi, la urechi – și la nas, piele etc. Diferențele dintre semnalele senzoriale prezise și cele reale dau naștere erorilor de predicție. În timp ce predicțiile perceptive se transmit predominant top-down (interior-spre-exterior), erorile de predicție se transmit bottom-up (exterior-spre-interior). Semnalele erorilor de predicție sunt folosite de către creier pentru a-și actualiza predicțiile în așa fel încât să fie pregătit de următoarea serie de input-uri senzoriale. Ceea ce percepem este rezultatul tuturor predicțiilor top-down luate împreună, după ce erorile predicțiilor senzoriale au fost minimizate – sau justificate – cât mai mult posibil.

Ipoteza halucinației controlate are multe trăsături comune cu alte teorii „predictive" ale percepției și funcționării cerebrale, în principal cu *procesarea predictivă*. Există, totuși, o diferență de accent. Procesarea predictivă este o teorie despre *mecanismele* prin care creierul realizează percepția (și cogniția și acțiunea). Halucinația controlată se referă, dimpotrivă, la cum anume explică mecanismele cerebrale *proprietățile fenomenologice* ale percepției conștiente. Cu alte cuvinte, procesarea predictivă este o teorie despre cum anume funcționează creierul, în timp ce halucinația controlată dezvoltă această teorie pentru a explica natura experiențelor conștiente. Este important că ambele se sprijină pe procesul minimizării erorii de predicție.

Minimizarea erorii de predicție este cea care face conexiunea dintre halucinația controlată și inferența baye-

Scanned with OKEN Scanner

siană. Preia ipoteza bayesiană despre *ce anume ar trebui* să facă creierul și o preschimbă într-o propunere despre ceea ce *chiar face*. Prin minimizarea erorilor de predicție peste tot și tot timpul, se pare că creierul implementează, de fapt, regula lui Bayes. Mai precis, *aproximează* regula lui Bayes. Este conexiunea care acreditează ideea conform căreia conținutul perceptiv este, mai degrabă, o halucinație controlată top-down decât o „recitare" bottom-up a datelor senzoriale.

Fie trei componente fundamentale ale minimizării erorii de predicție din creier: modelele generative, ierarhiile perceptive și „ponderarea preciziei" a semnalelor senzoriale.

Modelele generative determină repertoriul de lucruri perceptibile. Pentru a percepe o gorilă, creierul meu trebuie să fie echipat cu un model generativ capabil să genereze semnalele senzoriale relevante – acele semnale senzoriale asociate cu prezența reală a unei gorile. Aceste modele furnizează fluxul de predicții perceptive cu care sunt comparate datele senzoriale dinspre exterior pentru a formula erori de predicție – care, apoi, oferă predicții actualizate, pe măsură ce creierul încearcă să minimizeze aceste erori.

Predicțiile perceptive se întind pe multe scale de spațiu și timp, astfel că putem percepe o lume structurată plină de obiecte, oameni și locuri. O predicție de nivel înalt de a vedea o gorilă generează o predicție de nivel scăzut referitoare la membre, ochi, urechi și blană, la rândul lor generatoare de alte predicții despre culori, texturi și margini și, în final, în variații anticipate referitoare la strălucirile din câmpul vizual. Aceste ierarhii perceptive se suprapun peste simțuri și chiar pot depăși cu totul limitele senzorialului. Dacă aud brusc vocea mamei mele, cortexul meu vizual și-ar putea ajusta predicțiile presupunând că silueta care se apropie este a ei. Dacă sunt la zoo, regiunile perceptive

Scanned with OKEN Scanner

ale creierului meu vor fi mai pregătite să vadă gorila decât atunci când mă aflu pe stradă.

Este important să clarificăm aici faptul că „predicția" din minimizarea erorii de predicție nu se referă neapărat la viitor. Înseamnă pur și simplu a merge dincolo de date prin utilizarea unui model. În statistică, esența predicției este de a acoperi lipsa datelor. Nu contează dacă este vorba de o predicție pentru viitor – ne putem gândi la viitor ca fiind „date insuficiente" – sau de o situație din prezent, incomplet cunoscută.

Elementul cheie al minimizării erorii de predicție este ponderarea cu precizie. Am văzut deja cum anume fiabilitatea relativă a semnalelor senzoriale determină cât de mult sunt actualizate inferențele perceptive. Prima privire întâmplătoare către o gorilă aflată în depărtare sau peste pajiștea vecinului, printr-o fereastră murdară, va livra semnale senzoriale cu fiabilitate mică, astfel că prima ta cea mai bună estimare bayesiană nu se va deplasa prea mult. Am văzut și maniera în care fiabilitatea este surprinsă de precizia distribuției corespunzătoare de probabilități. Așa cum arată diagrama de la p. 107, datele senzoriale cu precizie estimată joasă au un efect mai slab asupra actualizării convingerilor precedente.

Folosesc mai degrabă „precizie estimată" decât pur și simplu „precizie", deoarece precizia semnalelor senzoriale nu este ceva furnizat direct creierului perceptiv. Trebuie și inferat. Creierul se confruntă nu numai cu sarcina de a depista cele mai plauzibile cauze ale input-urilor senzoriale, ci și cu verificarea fiabilității input-urilor senzoriale relevante. În practică, asta înseamnă că creierul ajustează permanent influența semnalelor senzoriale asupra inferenței perceptive. Face asta modificând tranzitoriu precizia lor estimată. La asta se referă termenul de „ponderare a preciziei". Sub-ponderarea preciziei estimate înseamnă că semnalele senzoriale au influență

Scanned with OKEN Scanner

mai mică asupra actualizării celei mai bune predicții, în timp ce supra-ponderarea ei înseamnă inversul: o influență mai puternică a semnalelor senzoriale asupra inferenței perceptive. Astfel, ponderarea preciziei joacă un rol esențial în coregrafia delicatului dans dintre predicții și erorile de predicție, necesară pentru a ajunge la cea mai bună estimare perceptivă.

Deși sună complicat, suntem cu toții foarte familiari cu rolul cântăririi preciziei în percepție. Creșterea preciziei estimate a semnalelor senzoriale nu înseamnă nimic altceva decât „a acorda atenție". Atunci când ești atent la ceva – de exemplu, atunci când chiar vrei să afli dacă e o gorilă acolo, în depărtare – creierul tău crește ponderea preciziei pentru semnalele senzoriale corespunzătoare, ceea ce e echivalent cu a crește fiabilitatea lor estimată sau cu reconsiderarea „ponderii" lor (turning up their „gain"). Atunci când abordăm atenția în acest mod putem înțelege de ce, uneori, nu vedem unele lucruri, chiar dacă acestea se află în fața noastră și chiar ne uităm la ele. Dacă ești atent la anumite semnale senzoriale – crescând, deci, precizia lor estimată – atunci alte semnale senzoriale vor avea influență mai mică în actualizarea celor mai bune estimări perceptive.

În mod remarcabil, în anumite situații, date senzoriale nesesizate pot să nu aibă deloc influență. În 1999, psihologul Daniel Simons a făcut o demonstrație a acestui fenomen într-un video bine-cunoscut, pe care l-a numit „orbire neintenționată". Dacă nu l-ați văzut deja, vizionați-l înainte să citiți mai departe.*

Iată despre ce e vorba. În demo, subiecții lui Simons privesc un film scurt, în care joacă două echipe de câte trei persoane. O echipă e în negru, cealaltă, în alb. Fiecare echipă are o minge de baschet care trece de la un membru al echipei la altul, în timp ce aceștia se deplasează oarecum la întâmplare. Treaba privitorului e să numere de câte ori

* www.youtube.com/watch?v=vJG698U2Mvo.

Scanned with OKEN Scanner

mingea trece de la un membru la altul al echipei albe. Este o sarcină care cere destul efort, deoarece sunt șase oameni care se plimbă peste tot și două mingi pasate permanent.

Uimitor este că, în timp ce numără, majoritatea nici nu observă persoana costumată în gorilă neagră care intră în scenă prin stânga, face tot felul de mișcări de gorilă, apoi iese prin dreapta. Arătându-le același video din nou, aceleași persoane sunt puse să observe gorila: o văd imediat și, de multe ori, nu le vine să creadă că era acolo de la bun început. Concentrarea atenției pe echipa în alb face ca semnalelor senzoriale de la echipa în negru – și de la gorilă – să li se aloce precizie estimată scăzută, ceea ce are influență foarte mică, spre deloc, asupra actualizării celei mai bune estimări perceptive.

Ceva asemănător mi s-a întâmplat într-o după-amiază, cu mulți ani în urmă, în timp ce conduceam spre locul preferat de surf din San Diego. Am făcut stânga acolo unde tocmai fusese instalat un semn cu „interzis stânga"; un drum secundar scurt care cobora către ocean, în vecinătatea Del Mar. Nu exista niciun motiv evident pentru acel semn, alte mașini din fața mea făcuseră și ele stânga, aceeași stângă pe care o făcusem de sute de ori de-a lungul anilor, așa că am fost foarte revoltat pentru amenda nedreaptă și am scris în depoziție că, practic, nici nu văzusem semnul, chiar dacă, „în principiu", era vizibil. Am invocat, în apărarea mea, principiile orbirii neintenționate – da, era un semn nou dar, din cauza cântăririi preciziei din creierul meu, nu am fost capabil să-l percep. Cazul a ajuns până la completul de judecată pentru trafic auto al Californiei, nu chiar la Curtea Supremă, dar suficient de sus ca să-mi văd numele pe lista „criminalilor zilei". Am pregătit inclusiv o scurtă prezentare în Power Point pentru judecător, care nu a fost de niciun folos.

Și magicienii se folosesc de orbirea neintenționată, chiar dacă nu-și descriu meșteșugul în astfel de termeni.

Scanned with OKEN Scanner

Trucurile detaliate, mai ales, cer un talent deosebit de a distrage atenția, astfel ca nimeni să nu remarce că dama de pică e strecurată în spatele urechii, de unde apare, apoi, ca din senin. Hoții de buzunare se bazează și ei pe acest neajuns al fiziologiei percepției. Am asistat odată la o astfel de scenă, în care Apollo Robbins, maestru al hoțiilor de buzunar, i-a ușurat de ceasuri, portofele și genți pe colegii mei – lucru cu atât mai remarcabil cu cât mulți dintre ei sunt experți în percepție, știu totul despre orbirea neintenționată și erau perfect conștienți de ceea ce va face Robbins.

E tentant să gândim despre interacțiunea noastră cu lumea în felul următor. Mai întâi, percepem lumea așa cum este ea. Apoi decidem ce facem. Apoi facem. *Simt, gândesc, acționez.* Așa pare că stau lucrurile, numai că, încă o dată, cum pare că stau lucrurile reprezintă un ghid foarte slab despre cum stau ele de fapt. E timpul să intre în scenă *acțiunea*.

Acțiunea nu poate fi separată de percepție. Percepția și acțiunea sunt cuplate atât de strâns, încât se determină și se definesc reciproc. Fiecare acțiune influențează percepția prin schimbarea datelor senzoriale care vin dinspre exterior, iar fiecare percepție este în felul în care este pentru a ajuta în ghidarea acțiunii. Percepția nu are rost, pur și simplu, în lipsa acțiunii. Percepem lumea din jurul nostru pentru a acționa eficient în acea lume, pentru a ne atinge scopurile și – pe termen lung – pentru a ne crește șansele de supraviețuire. Nu percepem lumea așa cum e ea, o percepem pentru că ne este util să o facem.

Se poate chiar ca acțiunea să fie prima. În loc să ne imaginăm creierul ca formulând cea mai bună estimare pentru a ghida comportamentul ulterior, putem să ni-l

Scanned with OKEN Scanner

imaginăm ca fiind ocupat în special cu generarea de acțiuni, pe care le calibrează continuu prin intermediul semnalelor senzoriale, astfel încât organismul să-și atingă scopurile. În această abordare, creierul este un sistem activ, deci dinamic, care extrage permanent probe din mediu, examinând consecințele*.

În procesarea predictivă, acțiunea și percepția sunt fețe ale aceleiași monede. Ambele sunt susținute de minimizarea erorilor de predicție senzorială. Până acum, am descris procesul de minimizare în termeni de actualizare a predicțiilor perceptive, numai că aceasta nu e singura variantă. Erorile de predicție pot fi reduse prin executarea de acțiuni dedicate schimbării datelor senzoriale, astfel ca noile date senzoriale să se potrivească unei predicții existente. Minimizarea erorii de predicție prin acțiune se numește *inferență activă* – termen definit de Karl Friston, specialist britanic în neuroștiință.

O manieră utilă de a gândi despre inferența activă este un fel de predicție perceptivă autosuficientă, un proces prin care creierul caută, executând diverse acțiuni, datele senzoriale care fac ca predicțiile lui perceptive să devină adevărate. Aceste acțiuni pot fi unele extrem de simple, cum ar fi mișcarea ochilor. În dimineața asta, îmi căutam cheile mașinii printre obiectele așezate peste tot pe birou. Odată cu mișcările ochilor, pe care îi îndreptam spre diverse puncte de pe birou, se actualizau predicțiile vizuale în fiecare moment (cană goală, cană goală, clame de hârtie, cană goală...), iar concentrarea mea vizuală interoga permanent tabloul privit, până când predicția perceptivă referitoare la cheile mașinii era îndeplinită.

* Ascidiile (nevertebrate marine) au, în stadiul juvenil, un fel de creier bine definit, dar rudimentar, pe care îl folosesc pentru a alege o piatră sau o ramură de coral pe care își vor petrece tot restul vieții, hrănindu-se cu ceea ce filtrează din curenții marini sau cu ce mai pică. Odată găsit locul potrivit, își digeră propriul creier, păstrându-și doar un sistem nervos foarte simplu. Unii au folosit exemplul ascidiilor ca exemplu pentru cariera academică – înainte și după găsirea unui loc permanent în universitate.

Scanned with OKEN Scanner

Orice tip de acțiune corporală va schimba datele senzoriale într-un fel sau altul, fie că vorbim de mișcarea ochilor, de deplasarea în altă cameră sau de contracția mușchilor abdominali. Chiar și „acțiuni" de nivel înalt, cum ar fi solicitarea unui nou loc de muncă sau decizia de a te căsători se descompun într-o succesiune de acțiuni corporale care afectează input-urile senzoriale. Fiecare tip de acțiune are potențialul de a suprima erorile de predicție senzorială prin inferență activă, astfel că fiecare tip de acțiune are rol direct în percepție.

La fel ca orice aspect al procesării predictive, inferența activă depinde de modelarea generativă. Mai specific, inferența activă se bazează pe abilitatea modelelor generative de a prezice consecințele senzoriale ale acțiunilor. Acestea sunt predicții de forma „dacă privesc într-acolo, ce date senzoriale aș putea primi?". Astfel de predicții se numesc *condiționale* – predicții despre ce anume *s-ar putea* întâmpla acolo unde s-ar pune problema. Fără predicții condiționale de acest tip, creierul nu are nicio posibilitate să știe care acțiune dintre nenumăratele posibile ar fi cea mai potrivită pentru a reduce erorile de predicție senzorială. Acțiunile despre care creierul meu prezice că ar fi cele mai potrivite pentru a-mi găsi cheile mașinii implică scanarea vizuală a biroului, nu privitul pe geam sau să-mi mișc mâna prin aer.

Pe lângă îndeplinirea predicțiilor perceptive existente, inferența activă poate ajuta și la îmbunătățirea respectivelor predicții. Pe perioade scurte de timp, acțiunile pot colecta noi date senzoriale pentru a ajuta la formularea de cele-mai-bune-estimări mai potrivite sau pentru a decide care dintre ipotezele perceptive e de preferat. Am văzut un astfel de exemplu la începutul acestui capitol, unde aveam de ales între ipoteza „ploii peste noapte" și cea a „aspersoarelor uitate deschise", iar alegerea variantei corecte era posibilă dacă priveam pajiștea vecinului. Un alt exemplu ar fi curățarea biroului de ceștile de cafea pentru

Scanned with OKEN Scanner

a-mi găsi cheile mașinii. În fiecare caz, alegerea acțiunii relevante se bazează pe un model generativ capabil să prezică modul în care datele senzoriale s-ar schimba ca rezultat al acelei acțiuni.

Pe termen lung, acțiunile sunt esențiale pentru *învățare* – care aici înseamnă îmbunătățirea modelelor generative ale creierului prin revelarea a cât mai mult despre cauzele semnalelor senzoriale și despre structura cauzală a lumii în general. Atunci când mă uit peste gard ca să pot infera motivele pentru care o anumită pajiște e udă, învăț mai multe despre motivele pentru care o pajiște poate fi udă în general. În cel mai bun caz, inferența activă generează un cerc virtuos în care acțiuni bine alese dezvăluie informație utilă despre structura lumii, care este, apoi, încorporată în modele generative îmbunătățite, care pot permite, apoi, inferență perceptivă îmbunătățită și noi acțiuni directe despre care se presupune că oferă și mai multă informație utilă.

Cel mai contraintuitiv aspect al inferenței active este acela că *acțiunea în sine* poate fi gândită ca o formă de predicție perceptivă autoîmplinită. Din acest punct de vedere, acțiunile nu doar participă la percepție – acțiunile *sunt* percepții. Atunci când îmi mișc ochii în căutarea cheilor mașinii sau când apuc cănile ca să fac curat, ceea ce se întâmplă este că predicțiile perceptive despre poziția și mișcarea corpului meu se adeveresc.

În inferența activă, acțiunile sunt predicții *proprioceptive* autoîmplinite. Propriocepția este o formă de percepție care înregistrează unde anume este poziționat corpul și cum anume se mișcă, prin colectarea semnalelor senzoriale de la receptorii specifici situați peste tot în oase și mușchi. Nu ne gândim prea mult la propriocepție deoarece este oarecum implicită, numai că simplul fapt că poți să-ți atingi nasul cu ochii închiși – încearcă! – demonstrează rolul fundamental pe care îl are în toate acțiunile noastre. Din perspectiva

Scanned with OKEN Scanner

inferenței active, atingerea nasului propriu înseamnă că e posibil ca o suită de predicții proprioceptive despre mișcarea mâinii și despre poziție să devină autoîmplinite – copleșind dovada senzorială că degetele mele *nu* ating nasul în mod obișnuit. Cântărirea preciziei joacă un rol important și aici. Pentru ca predicțiile proprioceptive să se adeverească, erorile de predicție care comunică creierului unde se află corpul, de fapt, trebuie atenuate sau scăzute/sub-ponderate. Este, astfel, vorba de un opus al acordării atenției – un fel de „neatenție" către corp, ceea ce-i permite să se miște.

Atunci când gândim în acest fel despre acțiune, subliniem faptul că acțiunea și percepția sunt cele două fețe ale unei monede. În loc să considerăm percepția ca fiind input-ul, iar acțiunea, output-ul în raport cu o minte „centrală", acțiunea și percepția sunt, amândouă, forme ale predicției bazate pe creier. Ambele depind de un proces comun de cea-mai-bună-estimare bayesiană, de coregrafia detaliată a dansului dintre predicțiile perceptive și erorile de predicție senzorială, cu diferențe între cine conduce și cine este condus.

Să revenim pentru ultima dată la creierul pe care ni l-am imaginat, ferecat în celula lui osoasă. Știm acum că acest creier nu e nici pe departe izolat. Înoată într-un torent de semnale senzoriale ce vin din afara și din interiorul corpului, direcționând permanent acțiuni – predicții proprioceptive autoîndeplinite – care sculptează proactiv fluxul senzorial. Barajul senzorial aferent este întâmpinat de o cascadă de predicții top-down, cu semnale de eroare de predicție trimise în sus pentru a stimula predicții din ce în ce mai bune și a declanșa noi acțiuni. Acest proces de rulare generează o aproximare a inferenței bayesiene, un

Scanned with OKEN Scanner

bayesianism-suficient-de-bun, în care creierul se fixează și refixează în timp ce evoluează spre cea-mai-bună-estimare a cauzelor mediului său senzorial. Astfel, se generează o lume perceptivă realistă – o halucinație controlată.

Înțelegând halucinațiile controlate în acest mod, avem, acum, motive suficiente pentru a recunoaște că predicțiile top-down nu doar că ne influențează percepțiile, ci chiar *sunt* ceea ce percepem. Lumea pe care o percepem, cu ale ei culori, forme și sunete nu este nimic mai mult, nici mai puțin decât cea-mai-bună-estimare a creierului nostru privind cauzele ascunse ale input-urilor sale senzoriale lipsite de culoare, de formă și de sunet.

Și, așa cum vom vedea în continuare, nu numai percepția pisicilor, a cănilor de cafea sau a gorilelor poate fi explicată astfel – ci *fiecare* aspect al experienței noastre perceptive.

Scanned with OKEN Scanner

6
Contribuția privitorului

Incursiunea noastră în structura profundă a experienței perceptive începe cu o excursie la Viena, la început de secol XX. Dacă v-ați fi aflat prin cafenelele și saloanele – pline de opium, pe vremea aceea – ale acestui oraș faimos, ați fi întâlnit mai multe personaje remarcabile. Era cercul de filosofi de la Viena, un grup ce-i includea pe Kurt Gödel, Rudolf Carnap și, din când în când, pe Ludwig Wittgenstein. Erau pionierii picturii moderniste, Gustav Klimt, Oskar Kokoschka și Egon Schiele, precum și istoricul de artă Alois Riegl. Și, bineînțeles, Sigmund Freud.

În atmosfera intelectuală fluidă din Viena de la acea vreme, cele două lumi, a artei și a științei, se amestecau într-o proporție neobișnuit de mare. Știința nu era considerată superioară artei, în sensul mult prea familiar în care arta și răspunsurile pe care le induce omului ar cere o explicație științifică. Nici arta nu se plasa undeva, dincolo de posibilitatea științei de a ajunge la ea. Artiștii și oamenii de știință – și criticii lor – erau aliați în încercarea de a înțelege experiența umană în întreaga ei bogăție și varietate. Nu degeaba Eric Kandel, specialistul în neuroștiință, a intitulat această perioadă „epoca introspecției" („the age of insight"), în cartea sa cu același nume.

Una dintre cele mai influente idei generate în epoca explorării interioare este cea a „contribuției privitorului", formulată inițial de Riegl și popularizată ulterior de una dintre personalitățile majore ale istoriei artei din secolul al XX-lea, Ernst Gombrich – născut la Viena în 1909.

Scanned with OKEN Scanner

Ideea lor a subliniat rolul observatorului – privitorul – în „completarea" imaginară a unei opere de artă. Contribuția privitorului este acea parte a experienței perceptive cu care vine cel care percepe și care nu se găsește în opera de artă – sau în lume – în sine.

Conceptul de contribuție a privitorului cere să fie conectat cu teoriile predictive ale percepției – precum teoria halucinației controlate. Așa cum spunea Kandel, „conștientizarea faptului că percepția privitorului implică o inferență top-down l-a convins pe Gombrich de faptul că nu există «privire inocentă»: orice percepție vizuală se bazează pe clasificarea conceptelor și pe interpretarea informației vizuale. Nu putem percepe ceva ce nu putem clasifica".

Pentru mine, contribuția privitorului este evidentă mai ales în prezența operelor de artă ale unor Claude Monet, Paul Cezanne și Camille Pissarro. Stând în fața uneia dintre capodoperele impresionismului – cum ar fi *Brumă la Ennery*, pictat în 1873, actualmente expus la Musée d'Orsay în Paris – sunt absorbit în altă lume. Unul dintre motivele pentru care astfel de picturi captivează este că lasă loc observatorului, sistemului său vizual, să-și facă propria lucrare de interpretare. În pictura lui Pissarro, „răzuielile de pe paletă... pe o pânză murdară" – cum afirma criticul Louis Leroy – reușesc să creeze impresia perceptivă a unui câmp acoperit de brumă aspră.

Peisajele impresioniste își doresc să scoată artistul din actul de a picta, de a reface „privirea inocentă" a lui Gombrich, conferind pânzei acele variații de luminozitate care sunt materia primă pentru inferența perceptivă, mai degrabă decât rezultatul acestui proces. Pentru asta, artistul trebuie să dezvolte și să pună în practică o înțelegere sofisticată a modului în care se manifestă aspectele subiective, fenomenologice ale vederii. Fiecare lucrare poate fi înțeleasă ca fiind un exercițiu de inginerie

Scanned with OKEN Scanner

inversă a sistemului vizual uman, de la input-ul senzorial, pe tot parcursul, până la experiența subiectivă. Picturile devin experimente în percepția predictivă și în natura experiențelor conștiente generate de aceste procese.

Picturi ca cea a lui Pissarro sunt mai mult decât ecouri sau presentimente ale unei științe a percepției, iar contribuția privitorului oferă mai mult decât o versiune din istoria artei despre minimizarea erorii de predicție. Ceea ce Gombrich și alții ca el scot la lumină este aprecierea profundă a naturii experiențiale, fenomenologice a percepției – apreciere care se pierde cu ușurință printre detaliile precedentelor, probabilităților și erorilor de predicție.

„Atunci când afirmăm că petele și pensulările de pe pânzele impresioniste „capătă brusc viață", ne referim la faptul că am fost îndrumați să alcătuim un peisaj din acești stropi de pigment". Gombrich surprinde, aici, ceva fundamental despre percepția conștientă, ceva ce se aplică, dincolo de artă, naturii experienței în general. Atunci când experimentăm lumea ca fiind „cu adevărat acolo", nu ne referim la o revelare pasivă a realității obiective, ci la o proiecție realistă și prezentă – ajung la lumea din afară dinspre creier.

În laborator, eforturile de a descifra căile prin care așteptările perceptive susțin experiența subiectivă încep cu experimente simple. Una dintre predicțiile experimentale este aceea că ar trebui să percepem mai repede și mai ușor lucrurile pe care le așteptăm, decât pe cele pe care nu le așteptăm. Cu câțiva ani în urmă, Yair Pinto – ulterior, cercetător postdoc sub îndrumarea mea, iar acum profesor-asistent la Universitatea din Amsterdam – a vrut să testeze această ipoteză, concentrându-se, așa cum facem adesea, pe experiența vizuală.

Scanned with OKEN Scanner

Yair a utilizat o configurație numită „suprimarea succesiunii de instantanee", în care fiecărui ochi îi sunt prezentate imagini diferite. Unui ochi i se arată o poză – o casă sau o față umană – iar celuilalt, o succesiune rapidă de tipare, cu model alungit. Atunci când creierul încearcă să fuzioneze cele două imagini într-o singură scenă, nu reușește, instantaneele cu modele alungite, în schimbare, tinzând să predomine și să fie, deci, ceea ce vede persoana respectivă în mod conștient. Percepția conștientă a pozei este suprimată de „succesiunea rapidă de instantanee" cu forme alungite. Experimentul nostru, reprezentat mai jos, a folosit o versiune a acestei metode în care contrastul dintre formele alungite a tot scăzut în timp, în timp ce cel al pozei a crescut. Asta înseamnă că, după cel mult câteva secunde, poza – a casei sau a feței umane – este vizualizată în mod conștient.

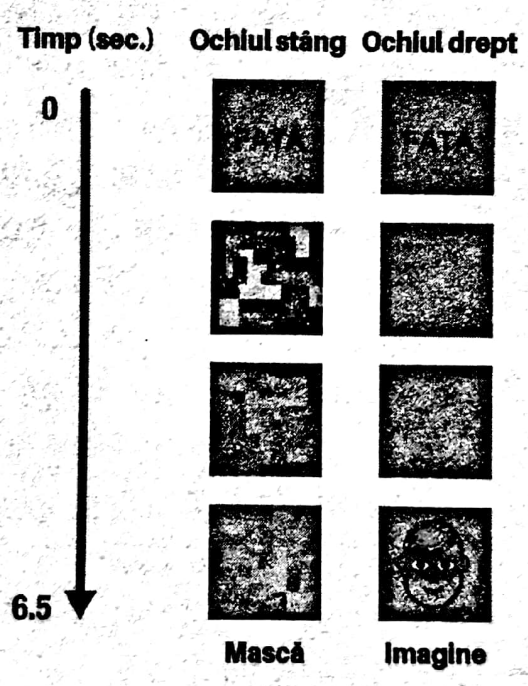

Fig. 10: Suprimarea succesiunii de instantanee cu contrast variabil.*

* Ochiului stâng i se prezintă un tipar în schimbare de forme alungite, al cărui contrast scade în timp, în timp ce ochiului drept i se prezintă o poză (o casă sau o față umană), al cărei contrast crește în timp. Pentru a direcționa fiecare imagine de la un monitor de computer la fiecare ochi se folosește un stereoscop cu oglindă. La începutul fiecărui experiment, participanților li se explică faptul că vor vedea o față sau o casă folosind cuvintele „față" și „casă" atunci când sunt arătate fiecărui ochi.

Scanned with OKEN Scanner

Pentru a afla cum anume sunt afectate percepțiile conștiente de către așteptările perceptive, am furnizat participanților, la începutul fiecărui experiment, fie cuvântul „casă", fie cuvântul „față". Important – aceste așteptări erau doar parțial valide. Atunci când participanților li se spunea cuvântul „față", li se arăta o față doar în 70% din cazuri, în restul de 30% arătându-li-se o casă. Pentru cuvântul „casă" am procedat invers. Măsurând timpul necesar pentru ca fiecare imagine din succesiunea de instantanee să devină sesizabilă, am putut determina cât de repede a fost vizualizată conștient o anumită imagine – casă sau față – atunci când subiecții se așteptau să o vadă, comparativ cu situația în care nu se așteptau.

Așa cum am prezis, subiecții vedeau mai repede și mai clar casa, atunci când se așteptau să o vadă – la fel și în cazul feței. Diferența de viteză a fost mică – aproximativ o zecime de secundă – dar fiabilă. În experimentul nostru, așteptările perceptive valide chiar duc la percepții conștiente mai rapide și mai clare.

Studiul nostru este unul dintre nenumăratele care examinează ce anume se întâmplă atunci când există așteptări perceptive. Într-un alt experiment elegant, Micha Heilbron, Floris de Lange și colegii lor de la Institutul Donders din Nijmegen au pornit de la așa-numitul „efect al superiorității cuvântului". Literele separate – cum ar fi U – sunt mai ușor de identificat atunci când fac parte dintr-un cuvânt, cum ar fi „CAUZĂ", decât dintr-o succesiune de litere care nu formează un cuvânt, cum ar fi „AEUVR". Echipa lui de Lange a arătat subiecților diferite exemple de cuvinte și de non-cuvinte, pe un fundal zgomotos. Experimentul lor a confirmat efectul superiorității cuvântului, literele luate separat fiind mai ușor de identificat într-un cuvânt decât într-un non-cuvânt.

Întorsătura, în acest studiu, a venit dintr-un mod inteligent de a analiza activitatea cerebrală a subiecților, cu

Scanned with OKEN Scanner

ajutorul RMNf. Prin analiza datelor cu ajutorul unei tehnici remarcabile numită „citire cerebrală“ (brain reading), s-a observat că semnătura neuronală a literei în cortexul vizual era mai „vizibilă“ – adică mai ușor de depistat comparativ cu reprezentarea neuronală a altor litere – atunci când litera era parte a unui cuvânt, decât dacă era parte a unui non-cuvânt. Asta înseamnă că așteptările perceptive oferite de contextul cuvântului pot afecta procesarea vizuală în stadiile sale incipiente într-o manieră care îmbunătățește percepția, exact așa cum sugerează ideea de halucinație controlată că ar trebui să se întâmple.

Oricât de revelatoare ar fi aceste experimente, mediile din laborator duc mare lipsă de bogăția și varietatea experiențelor conștiente din afara laboratorului. Pentru a ieși din laborator, e nevoie să gândim altfel.

Într-o zi de vară, nu cu mult timp în urmă și pentru prima dată în viață, mi-am pus un pic de LSD sub limbă și m-am întins în iarbă, să văd ce se întâmplă. Era o zi caldă, cu o briză blândă și un cer albastru-pal, cu câțiva nori împrăștiați. După aproximativ o jumătate de oră, exact cum i s-a întâmplat lui Albert Hofmann cu mulți ani înainte, totul a început să se deplaseze și să se schimbe. Dealurile și cerul și norii și marea au început să pulseze, devenind mai animate, aproape vii, captivându-mi profund, împletindu-se și confundându-se cu corpul meu. Ca orice om de știință adevărat, încercam să și notez, dar, privind ce am consemnat a doua zi, mi-am dat seama că încercările mele au fost sortite eșecului destul de repede. Îmi aminteam că norii luau forme diferite, dar clare, într-un fel care părea, cel puțin parțial, sub control. Odată ce un nor anume începea să se asambleze sub forma unui cal – sau a unei pisici sau a unei persoane – am descoperit că puteam accentua efectul final,

Scanned with OKEN Scanner

fără prea mult efort, uneori până la absurd. La un moment dat, la orizont a apărut o paradă a Cillei Black.*

Celor care nu sunt convinși că creierul este organul experienței, halucinațiile induse de LSD le pot oferi dovezi de netăgăduit. Câteva zile după, încă mai aveam impresia că pot vedea „prin" experiențele mele perceptive, trăindu-le – măcar parțial – sub forma constructelor care sunt. Am continuat să trăiesc ecouri ale mediului, împreună cu mesajul perceptiv.

Desigur, poți vedea fețe în nori și fără un stimulent farmaceutic – măcar idei sau sugestii de fețe, umbrele lor probabilistice proiectate pe și în cer. Fenomenul prin care vedem tipare peste tot se numește *pareidolie* (din παρά, «pe lângă, alături, în loc de» și εἶδόν, «imagine, contur, formă»). Pentru oameni – și câteva animale – semnificația fețelor este dovada că creierul este pre-încărcat cu așteptări precedente legate de fețe. De aceea avem cu toții tendința să vedem fețe peste tot – nori, felii de pâine, chiuvete – ca în imaginea de mai jos. Și pentru că toți facem asta, atunci nu ne gândim că pareidolia este o halucinație. Atunci când

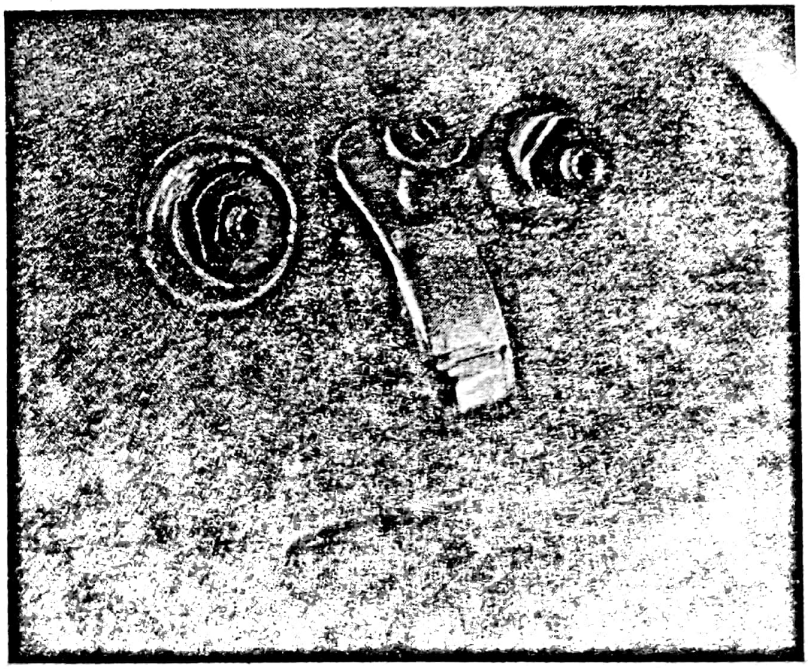

Fig. 11: Fața din chiuvetă.

* Cilla Black a fost pop star al anilor 60 și, ulterior, vedetă TV din Liverpool.

Scanned with OKEN Scanner

un schizofrenic aude o voce care-i poruncește să-și facă rău sau care-i spune că el e Iisus reîntrupat, dar nimeni altcineva nu mai aude acea voce, o numim halucinație. Atunci când, din cauza LSD, o văd pe Cilla Black mărșăluind pe cer, tot halucinație e.

Așa cum știm acum, ar fi o greșeală să gândim despre oricare dintre aceste fenomene – oricât de bizare ar fi ele – că sunt total diferite de obișnuita cea-mai-bună-estimare perceptivă. *Toate* experiențele noastre, că le considerăm halucinatorii sau nu, sunt mereu și oriunde fondate pe o proiecție a așteptărilor perceptive asupra și în mediul nostru senzorial. Ceea ce numim „halucinație" este ceea ce se întâmplă atunci când precedentele perceptive sunt deosebit de puternice, copleșind datele senzoriale în așa măsură încât convingerea creierului despre cauzele lor începe să se destrame.

Inspirați de această continuitate între percepția normală și halucinație, explorăm, în laboratorul nostru, noi căi de studiu a manierei în care cea-mai-bună-estimare perceptivă dă naștere experienței perceptive, iar experimentele realizate ne-au dus prin locuri ciudate.

Pornind din biroul meu, dacă urci 2 etaje și te strecori prin mațele vechiului departament de chimie, vei găsi unul dintre spațiile improvizate ale laboratoarelor noastre – a cărui localizare și scop sunt dezvăluite de o bucată de hârtie lipită pe ușă cu pastă adezivă: „VR/AR lab". Aici utilizăm tehnologiile cu dezvoltare rapidă ale realității virtuale și augmentate (VR/AR) pentru a investiga percepția lumii și a sinelui, în maniere altfel imposibile. Acum câțiva ani, am decis să construim o „mașină de halucinat", să vedem dacă putem genera experiențe asemănătoare halucinației într-o manieră controlabilă experimental, prin simularea precedentelor perceptive hiperactive. Proiectul a fost condus de Keisuke Suzuki, un cercetător senior postdoc din laboratorul nostru și expertul în VR.

Scanned with OKEN Scanner

Cu ajutorul unei camere video de 360 de grade, am înregistrat, mai întâi, o vedere panoramică a unui mediu din lumea reală. Am ales scuarul principal din campusul universitar, marți la prânz, atunci când studenții și personalul mișună în jurul tarabelor cu mâncare aduse acolo săptămânal. Am procesat apoi filmarea cu un algoritm scris de Keisuke, bazat pe procedura Google intitulată „deep dream" (vis adânc) pentru a genera o halucinație simulată.

Algoritmul „deep dream" implică preluarea unei rețele neuronale artificiale, care a fost antrenată să recunoască obiectele din imagini, și rularea ei în sens invers. Astfel de rețele sunt alcătuite din mai multe straturi de neuroni simulați, cu conexiunile aranjate în așa fel încât să semene, oarecum, cu calea bottom-up din sistemul vizual. Deoarece aceste rețele conțin doar conexiuni botom-up, sunt ușor de antrenat prin intermediul metodelor standard de învățare automată (machine-learning). Rețeaua pe care am folosit-o a fost antrenată să identifice mai mult de o mie de tipuri diferite de obiecte din imagini, inclusiv multe rase de câini. Face o treabă excelentă, diferențiind chiar și între varietățile de husky - care, mie, îmi par toate la fel.

Modul standard de utilizare a acestor rețele este de a le arăta o imagine și de a întreba, apoi, ce „crede" rețeaua că se află în imagine. Algoritmul deep-dream inversează procedeul, spunând rețelei că acolo există un anume obiect și actualizând, în schimb, obiectul. Cu alte cuvinte, algoritmul proiectează o predicție perceptivă asupra și într-o imagine, dându-i un supliment de contribuția privitorului. Pentru mașina de halucinat, am aplicat acest proces cadru cu cadrul întregii panorame video și am adăugat câteva zorzoane pentru a rezolva continuitatea imaginii șamd. Am reluat filmul deep-dreamed printr-un display așezat pe cap, astfel ca subiecții să poată privi în jur și să-l experimenteze într-un mod imersiv. Așa a luat naștere mașina de halucinat.

Scanned with OKEN Scanner

Când am probat-o prima dată, experiența a fost mult mai convingătoare decât mă așteptam. Departe de una generată de acid (LSD, n. trad.) sau de halucinația psihotică (atât cât știu), totul a fost, fără îndoială, substanțial transformat. Nu era nicio Cilla Black, de data asta câini și părți de câine înmugureau organic din toate părțile cadrului în care mă aflam, într-o manieră care părea total diferită de cea în care ai prelua și monta poze de câini într-un film preexistent (vezi imaginea de mai jos pentru un cadru alb-negru). Puterea mașinii de halucinat stă în capacitatea ei de a simula efectele celor-mai-bune-estimări top-down despre prezența câinilor, recapitulând, astfel, într-un mod exagerat, procesul prin care noi percepem și interpretăm scenele vizuale ale lumii reale.

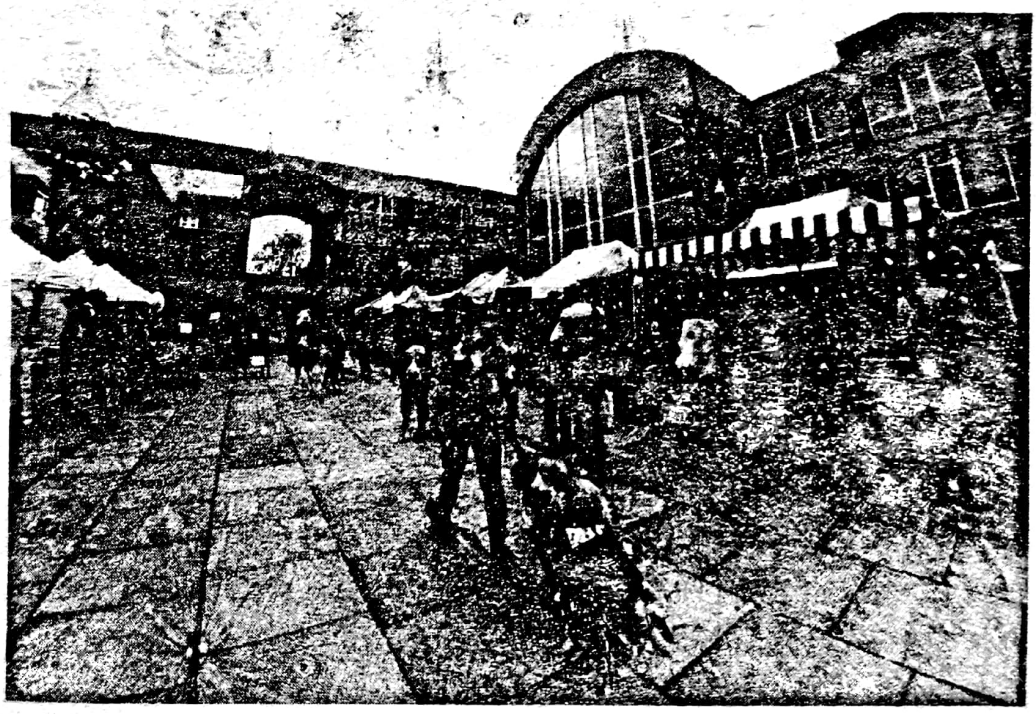

Fig. 12: Cadru din video-ul generat de mașina de halucinat.

Programând mașina de halucinat în diverse moduri, ușor diferite, putem genera diferite tipuri de experiențe halucinatorii simulate. De exemplu, dacă fixăm activitatea într-unul dintre straturile din mijloc ale rețelei – nu în stratul de output – obținem halucinații cu părți ale obiectelor, nu cu obiecte întregi. În acest caz, scena la care privești se umple

Scanned with OKEN Scanner

de ochi, urechi și labe, un amestec fără sens de părți de câine care-ți invadează lumea vizuală. Dacă ne fixăm în straturi și mai joase, ajungem să generăm ceva ce ar putea fi cel mai bine descris ca halucinații „geometrice", în care caracteristicile primare ale mediului vizual – margini, linii, texturi, tipare – devin neobișnuit de animate și de proeminente.

Mașina de halucinat este un exercițiu din ceea ce am putea numi „fenomenologie computațională": folosirea modelelor computaționale pentru a construi punți explicative între mecanismele și proprietățile experienței perceptive. Valoarea sa imediată stă în potrivirea arhitecturii computaționale a percepției predictive cu fenomenologia halucinației. Astfel, putem începe să înțelegem de ce anumite tipuri de halucinație sunt așa cum sunt. Dincolo de această aplicație se află, însă, ideea mult mai profundă, mult mai interesantă pentru mine, conform căreia, încercând să elucidăm misterele halucinației, putem înțelege mai bine și experiența perceptivă normală, de fiecare zi. Mașina de halucinat arată limpede, într-un mod personal, imediat și animat, că ceea ce numim halucinație este o formă de percepție necontrolată. Și că percepția normală – aici și acum – este, într-adevăr, o formă de halucinație controlată.

Ne-am putea îngrijora că ipoteza halucinației controlate s-ar limita la explicarea unor lucruri precum: „văd o masă pentru că asta e estimarea cea mai bună a creierului meu privind cauzele input-ului senzorial la acest moment" (sau, în loc de masă: față, pisică, câine, scaun roșu, cumnat, avocado, Cilla Black.) Cred că putem merge mult mai departe, pentru a explica ceea ce mie îmi place să numesc „structura profundă/deep structure" a percepției – modalitățile în care conținuturile conștienței apar în experiența noastră, în timp și spațiu și în diferite variante.

Scanned with OKEN Scanner

Să pornim de la observația banală că lumea noastră vizuală este alcătuită în mare din obiecte și din spațiile dintre ele. Când privesc la cana de cafea de pe birou, percep, într-un fel, și spatele ei, chiar dacă nu văd direct acea parte. Cana îmi pare că ocupă un volum anume, în timp ce o cană dintr-o fotografie sau dintr-un desen, nu. Aceasta este fenomenologia „obiectualității". Obiectualitatea este mai degrabă o proprietate a modului în care apar, la modul general, conținuturile vizuale, decât o proprietate a oricărei experiențe conștiente.

Deși obiectualitatea este o proprietate omniprezentă a experienței vizuale, nu este și universală. Dacă privești cerul de un albastru egal distribuit într-o zi senină, nu ai impresia că s-ar afla „acolo, undeva". Iar dacă privești direct la soare, apoi îți muți privirea, imaginea succesivă retiniană arsă în sistemul tău vizual este experimentată nu ca un obiect, ci ca o eroare temporară. Astfel de diferențe se observă și în alte situații: persoane cu tinnitus nu simt sunetele neplăcute ca având legătură cu lucruri reale din afară, de aceea această condiție este supranumită „țiuituri în ureche".

Artiștii au recunoscut de mult relevanța obiectualității pentru percepția umană. Ubicua *Trădare a imaginilor*, de René Magritte (pagina următoare), explorează diferența dintre un obiect și imaginea obiectului. O mare parte a operei cubistului Pablo Picasso investighează cum anume percepția obiectualului depinde de perspectiva noastră la persoana întâi. Picturile sale desfac și rearanjează obiecte în nenumărate feluri, reprezentându-le din mai multe perspective deodată. Putem considera despre aceste picturi, ca și despre altele la fel, că explorează principiile obiectualului din perspectiva contribuției privitorului. Opera lui Picasso, mai ales, atrage observatorul spre a crea în mod imaginativ obiecte perceptive dintr-un amestec de posibilități. Așa cum formulează filosoful Maurice Merleau-Ponty, artistul investighează, prin pictură, mijloacele prin care un obiect se face vizibil ochilor noștri.

Scanned with OKEN Scanner

În știința cognitivă, fenomenologia obiectualității a fost explorată cel mai amănunțit de „teoria contingenței senzorio-motorii". Conform acestei teorii, ceea ce trăim depinde de „stăpânirea practică" a modului în care acțiunile schimbă input-urile senzoriale. Atunci când percepem ceva, conținutul a ceea ce percepem nu este purtat de semnalele senzoriale; în schimb, provine din cunoașterea implicită a creierului despre cum anume sunt cuplate acțiunile și senzațiile. Din această perspectivă, vederea – și toate modalitățile noastre perceptive – sunt ceea ce *face* un organism, nu flux de informație pasivă către o „minte" centralizată.

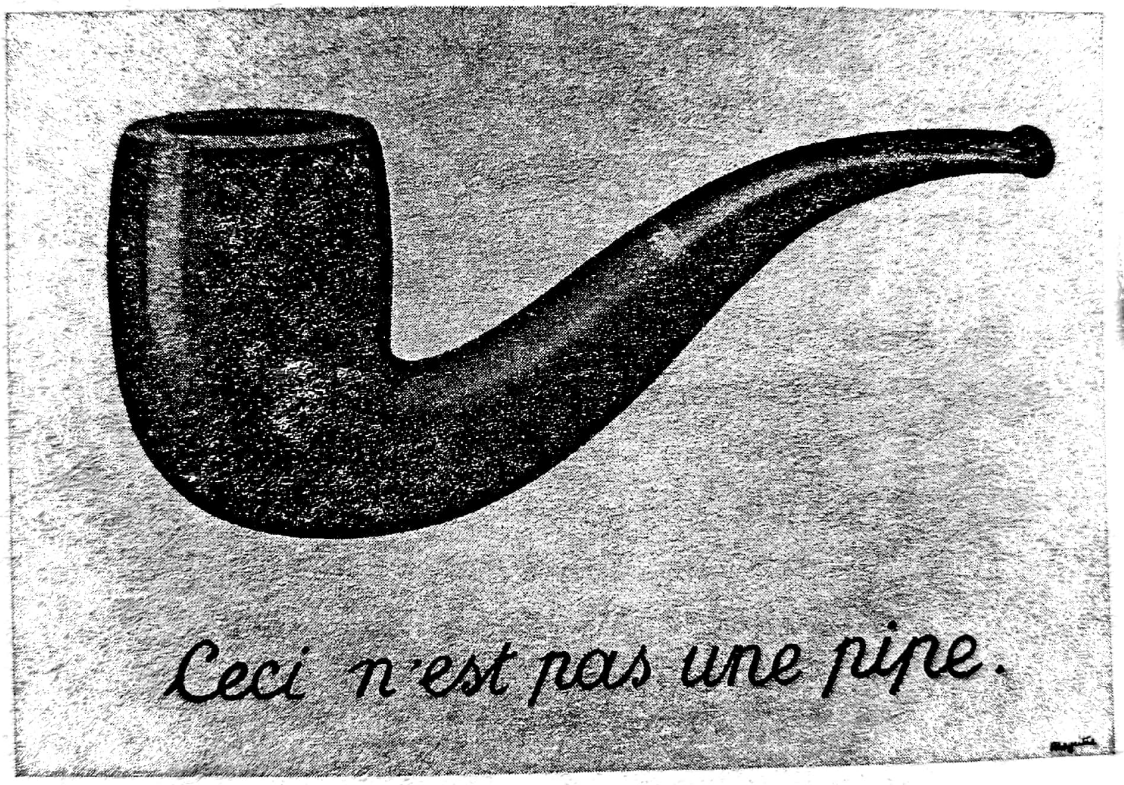

Fig. 13: René Magritte, *Trădarea imaginilor* (1929)
ADAGP, Paris & DACS, London 2021

În capitolul 4, am descris experiența roșului în termeni de predicții pe baze cerebrale despre cum anume suprafețele reflectă lumina. Să extindem acum această explicație la obiectualitate. Dacă țin o roșie în fața mea, percep că roșia *are un spate*, în fix același mod în care nu se întâmplă asta atunci când privesc o poză a roșiei (sau

Scanned with OKEN Scanner

o poză a unei pipe, ca în pictura lui Magritte) sau un cer albastru senin sau când am o post-imagine retiniană. Conform teoriei contingenței senzorio-motorii, devin conștient perceptiv de spatele roșiei chiar dacă nu îl pot vedea direct, ca urmare a cunoașterii implicite, bobinate în creierul meu, despre faptul că o roșie, când o răsucesc, schimbă semnalele senzoriale aferente.

Bobinarea necesară vine sub forma modelului generativ. Așa cum știm din capitolul anterior, modelele generative pot prezice consecințele senzoriale ale acțiunilor. Aceste predicții sunt „condiționale" sau „contrafactuale", în sensul că se referă la ceea ce *ar putea să se întâmple* sau la ceea ce *ar fi putut să se întâmple* semnalelor senzoriale, în condițiile anumitor acțiuni specifice. Într-un studiu pe care l-am scris acum câțiva ani, am propus ideea că fenomenologia obiectualității depinde de *bogăția* acestor predicții condiționale sau contrafactuale. Modelele generative care codifică multe predicții diferite de acest tip, cum ar fi aceea ce roșia are pielița roșie peste tot, trimit la o fenomenologie accentuată a obiectualului. Dar modelele generative care codifică doar câteva sau nicio astfel de predicție, cum e cazul cerului albastru, fără alte trăsături, sau o post-imagine retiniană vor conduce la obiectualitate scăzută sau absentă.

O altă situație în care obiectualitatea lipsește, de obicei, este cea a „sinesteziei grafem-culoare". Termenul de sinestezie se referă la un fel de „amestec al simțurilor". Cei cu varietatea grafem-culoare văd și culori odată cu literele: de exemplu, litera A produce senzația de roșu luminos, indiferent de culoarea pe care o are în realitate, pe pagină. Deși aceste culori apar clar și automat – ceea ce înseamnă că ori de câte ori subiectul vede o anumită literă, vede și aceeași culoare, fără a fi nevoie de un efort conștient în acest sens – sinesteții nu confundă culorile pe care le văd cu cele adevărate, „din lumea reală". Cred că asta se întâmplă deoarece culorile sinestetice, atunci când sunt comparate

Scanned with OKEN Scanner

cu cele „reale", nu suportă un repertoriu bogat de predicții senzorio-motorii. Un „roșu" sinestezic nu se schimbă prea mult pe măsură ce te deplasezi sau lumina ambientală se schimbă, așa că nu va exista nicio fenomenologie a obiectualității.

În laboratorul nostru VR, am început să testăm aceste idei. Într-un experiment recent, am creat o serie de obiecte virtuale special nefamiliare, fiecare definit de diferite vezicule și protruzii, pe care participanții le vedeau prin intermediul unui display montat pe cap (vezi mai jos). Am adaptat metoda suprimării instantaneelor, folosită în experimentul cu fețe/case, în așa fel încât fiecare obiect să fie inițial invizibil dar, în final, să ajungă conștientizat. Dacă în experimentul față/casă am manipulat imaginea să fie așteptată sau nu, instalația noastră de VR ne-a permis să manipulăm validitatea predicțiilor senzorio-motorii prin schimbarea manierei în care obiectele au răspuns la acțiuni. Participanții au folosit un joystick pentru a roti obiectele virtuale, iar noi le-am putut face să răspundă fie la fel ca un obiect real, fie să se deplaseze la întâmplare, impredictibil. Am prezis că obiectele virtuale care se comportă normal vor ajunge să fie conștientizate mai repede decât cele care au încălcat predicțiile senzorio-motorii și chiar așa a fost. Acest experiment este, evident, imperfect deoarece consideră viteza de acces la percepția conștientă ca fiind un proxy pentru fenomenologia obiectualului. Dar arată, totuși, că validitatea predicțiilor senzorio-motorii poate afecta percepția conștientă într-o manieră specifică și măsurabilă.

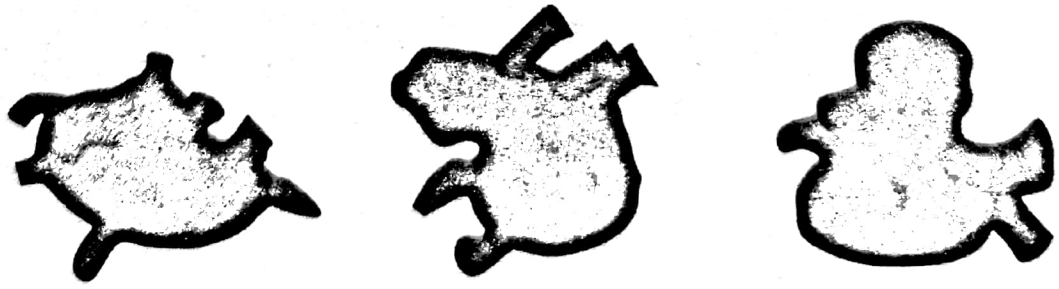

Fig. 14: Câteva obiecte virtuale, proiectate să arate neobișnuit.

Scanned with OKEN Scanner

Printre multele idei intuitive, dar false despre percepție este una conform căreia schimbările care survin în ceea ce percepem corespund direct schimbărilor exterioare. Dar schimbarea, ca și obiectualitatea, este o altă manifestare a structurii profunde a experienței perceptive. Schimbarea în percepție nu apare doar prin schimbarea datelor senzoriale. Percepem schimbarea prin aceleași principii de cea-mai-bună-estimare care generează toate celelalte aspecte ale percepției.

Multe experimente au arătat că schimbarea fizică – din lumea exterioară – nu este nici necesară, nici suficientă pentru percepția schimbării. Imaginea cu șerpi de mai jos oferă un exemplu foarte clar în care, deși nimic nu se mișcă, impresia perceptivă e de mișcare, mai ales când privirea se deplasează pe deasupra imaginii. Explicația este că detaliile

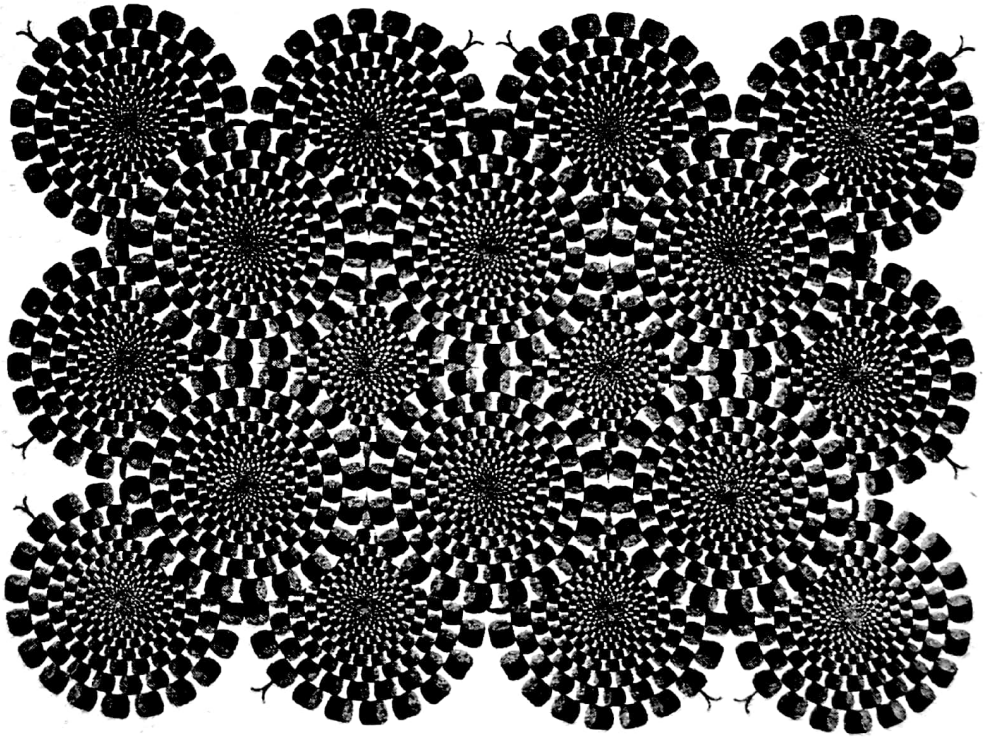

Fig. 15: Iluzia „șerpilor rotitori"
Credit: Akiyoshi Kitaoka

* Pentru versiunea colorată (mai eficientă) a acestei iluzii, vizitează www.illusionsindex.org/i/rotating-snakes.

Scanned with OKEN Scanner

fine ale imaginii, percepute în câmpul vederii periferice – situat dincolo de coada ochiului – conving cortexul vizual să infere mișcare, chiar dacă mișcarea lipsește.

Situația inversă – schimbare fizică fără schimbare perceptivă – are loc în „orbirea față de schimbare". Aceasta se întâmplă atunci când anumite aspecte ale mediului se schimbă foarte încet sau când totul se schimbă dintr-o dată, doar câteva caracteristici rămânând relevante. Într-o înregistrare remarcabilă a acestui fenomen,* jumătatea de jos a unei imagini își poate schimba complet culoarea – de la roșu la violet – dar, deoarece schimbarea are loc foarte încet, în peste 40 de secunde, majoritatea nu observă acest lucru, chiar dacă privesc direct la zona în schimbare. (Este obligatoriu, însă, ca subiecții să nu fie avertizați de schimbarea culorii. În caz contrar, sesizează schimbarea cu ușurință.) Acest exemplu are asemănări cu orbirea neintenționată, pe care am descris-o în capitolul anterior, unde subiecții nu sesizează prezența neașteptată a unei gorile în mijlocul unui joc de baschet. Diferența, aici, este că ceea ce nu observă oamenii este „schimbarea în sine".

Unii cred că orbirea față de schimbare expune o dilemă filosofică: după ce imaginea și-a schimbat culoarea, continui să trăiești roșul (chiar dacă acum e mov) sau trăiești movul, caz în care te poți întreba ce anume ai trăit înainte, dat fiind că nu ai trăit nicio schimbare? Rezoluția e să negi premisa întrebării și să recunoști că *percepția schimbării* nu e totuna cu *schimbarea percepției*. Trăirea schimbării este o altă inferență perceptivă, o altă variantă de halucinație controlată.

Iar dacă trăirile schimbării sunt inferențe perceptive, tot așa sunt și trăirile timpului.

---

* Vezi https://www.youtube.com/watch?v=hhXZng6o6Dk.

Scanned with OKEN Scanner

Timpul este unul dintre conceptele cele mai greu de înțeles atât în filosofie, cât și în fizică, dar și în neuroștiință. Fizica se străduiește să înțeleagă ce anume este și de ce curge (dacă chiar curge), iar provocarea pentru specialiștii în neuroștiință este la fel de spinoasă. Toate experiențele noastre perceptive au loc în și prin timp. Chiar și trăirea prezentului pare cumva distribuită între un trecut relativ fix și un viitor parțial deschis. Timpul curge și pentru noi, uneori târându-se, alteori, zburând.

Trăim secunde, ore, luni și ani, dar nu avem „senzori pentru timp" în creier. Pentru vedere, avem fotoreceptorii retinieni; pentru auz, există celule ciliare în ureche; nu există, însă, un sistem senzorial pentru timp. Mai mult, cu excepția ritmului circadian, din cauza căruia suferim de schimbarea de fus orar, printre altele, nu avem nicio dovadă că ar exista un „ceas neuronal" în capul nostru, care să ne măsoare experiențele în timp – și care ar fi, oricum, un prim exemplu a ceea ce Daniel Dennett a numit dublă transducție, în care o proprietate a lumii este re-instanțiată în creier în beneficiul unui presupus observator intern. În schimb, ca și schimbarea, ca și toate celelalte percepții ale noastre, trăirea timpului este tot halucinație controlată.

Controlată de ce anume, totuși? Fără un canal dedicat de transmitere, ce anume ar putea oferi echivalentul erorilor de predicție senzorială? O soluție simplă și elegantă a fost propusă de colegul meu, specialistul în științe cognitive Warrick Roseboom, care s-a alăturat Centrului nostru în 2015 și care acum își conduce propriul grup de cercetare pe tema percepției timpului. Ideea lui e că noi inferăm timpul pe baza vitezei de schimbare a conținuturilor perceptive în alt mod, nu a ticăitului unui ceas intern - și a configurat o metodă inteligentă de a o testa.

Sub conducerea lui Warrick, echipa noastră a făcut o bibliotecă de înregistrări video de diferite durate și în contexte diferite – o stradă urbană aglomerată, un birou

Scanned with OKEN Scanner

gol, câteva vaci păscând pe câmpul de lângă universitate. Apoi am apelat la voluntari care au privit acele înregistrări și au estimat durata fiecăruia. Aici au ieșit la iveală prejudecățile caracteristice: subestimarea duratei filmelor lungi, supraestimarea duratei celor scurte. De asemenea, prejudecăți legate de contextul fiecărei scene, cele aglomerate fiind percepute ca mai lungi decât cele liniștite, chiar și pentru filme care aveau clar aceeași durată.

Warrick a prezentat aceleași înregistrării video unei rețele neuronale artificiale care a mimat funcționarea sistemului vizual uman. Această rețea era, de fapt, aceeași pe care am folosit-o în mașina de halucinat. Pentru fiecare video, s-a calculat o estimare a duratei pe baza – aproximativ – vitezei de schimbare acumulate a activității în cadrul rețelei. Aceste estimate nu au implicat niciun fel de „ceas intern". Remarcabil, estimatele rețelei neuronale și estimatele umane au fost, practic, identice, demonstrând aceleași prejudecăți legate de durată și de context. Asta arată că percepția timpului poate apărea, măcar în principiu, dintr-o „cea mai bună estimare" a vitezei de schimbare a semnalelor senzoriale, fără să fie nevoie de un pacemaker intern.

Recent am dus această cercetare mai departe căutând dovezi ale acestui proces în creier. Într-un studiu condus de cercetătorul postdoc Maxine Sherman, am folosit RMNf pentru a înregistra activitatea cerebrală a subiecților în timp ce priveau același set de înregistrări video și estimau durata fiecăruia. Am vrut să știm dacă putem utiliza activitatea din cortexul vizual pentru a prezice cât de lung pare un video, așa cum făcusem în studiul anterior al lui Warrick, cu ajutorul modelului computațional al vederii. Maxine a arătat că putem. Activitatea cerebrală din sistemul vizual, dar nu și din alte zone ale creierului, prezicea cu acuratețe durata subiectivă. Aceasta este o dovadă de netăgăduit că experiența duratei chiar apare din cea-mai-bună-estimare perceptivă, nu din ticăitul vreunui ceas neuronal.

Scanned with OKEN Scanner

Alte experimente care ar fi vrut să demonstreze existența unui „ceas intern" au eșuat. Preferatul meu, de departe, implică săritul în gol de pe macarale. Specialistul în neuroștiință David Eagleman a abordat ideea generală conform căreia timpul subiectiv încetinește în momente dramatice, cum ar fi cele de dinaintea unui accident de mașină. Această încetinire subiectivă s-ar putea datora, spune el, unui ceas intern care ar merge mai repede – cu cât mai multe ticăituri ale ceasului într-o anumită perioadă de timp, cu atât mai lungă durata percepută. Asta ar face ca, în schimb, viteza de percepție să fie mai mare, deoarece un ceas mai rapid ar însemna o capacitate îmbunătățită de a percepe durate scurte.

Pentru a testa această idee, Eagleman și echipa sa au proiectat un ceas digital special, care afișează mai multe numere ce se succed, însă, atât de repede încât devin imposibil de citit în condiții normale. Voluntarii care s-au oferit să participe la acest studiu mai aveau ceva de făcut: să privească la respectivele ceasuri în timp ce săreau în gol – timp în care, evident, nivelul adrenalinei creștea. Dacă ceasul intern și-ar mări viteza, atunci – după logica lui Eagleman – voluntarul ar fi devenit capabil să identifice numerele din imaginea cețoasă generată de ceasul special, în timpul căderii libere. Or, nu s-a întâmplat asta, așa că studiul nu a putut demonstra existența unui ceas intern. Evident, lipsa dovezii nu este totuna cu dovada lipsei, dar... ce experiment!

La granițele cercetării noastre în structura profundă a experienței perceptive se află un proiect care investighează percepția „realității" în sine. Acest proiect, condus de Alberto Mariola, un doctorand talentat din laborator, implică un nou aranjament experimental pe care îl numim

Scanned with OKEN Scanner

„realitate substitutivă". Oricât de imersive ar fi, mediile VR se deosebesc clar de lumea reală. Voluntarii din mașina noastră de halucinat știau întotdeauna că ceea ce trăiesc nu este real, oricât de razna ar lua-o toate. Realitatea substitutivă tinde să depășească această limitare. Scopul este de a crea un sistem în care oamenii să trăiască mediile în care se află ca fiind reale și să creadă că sunt reale, chiar dacă nu este așa.

Ideea e simplă. Ca și în cazul mașinii de halucinat, am înregistrat video mai multe panoramări, de data asta din interiorul laboratorului VR/AR în care ne desfășurăm experimentele. Când voluntarii ajung în laborator, se așază pe un taburet în mijlocul camerei și își pun pe cap dispozitivul VR care are o cameră atașată în față. Sunt invitați să privească în jur, prin respectiva cameră. La un moment dat schimbăm sursa, astfel încât camera trimite imagini nu din lumea reală, ci din înregistrarea video. Remarcabil, majoritatea celor care trec prin această experiență continuă să creadă că ceea ce văd e din „realitate", chiar dacă nu mai e.

Cu acest aranjament, putem testa acum ipoteze despre condițiile în care oamenii percep mediul în care se află ca fiind real și – poate mai important – de ce anume e nevoie pentru ca acest aspect, de obicei omniprezent, al experienței conștiente, să se destrame. Astfel de situații se pot întâmpla, și chiar se întâmplă, nu numai în cazuri precum post-imaginile retiniene, ci și în tulburările psihiatrice debilitante precum depersonalizarea sau derealizarea, cazuri în care poate avea loc o pierdere globală a realității trăite a lumii, dar și a sinelui.

Investigarea a ce anume face ca o experiență perceptivă să pară „reală" ne duce înapoi la analiza lui Wittgenstein despre revoluția copernicană: aceea că chiar dacă înțelegem cum anume *sunt* lucrurile – Terra rotindu-se în jurul Soarelui, percepția ca halucinație controlată – în multe

Scanned with OKEN Scanner

alte aspecte, aceleași lucruri *par* să fie, de fapt, la fel ca la bun început. Când privesc la scaunul roșu din colțul camerei, roșul său (și condiția lui de scaun) încă *pare să existe* cu adevărat – veridic –, proprietate, mai degrabă, a unei realități independente de minte, decât o construcție elaborată a unui creier care face cea-mai-bună-estimare.

Cu mult timp în urmă, în secolul al XVIII-lea, David Hume a făcut o observație similară despre cauzalitate – o altă caracteristică omniprezentă a felului în care trăim lumea. În loc să considere cauzalitatea fizică drept o proprietate obiectivă a lumii, imediat accesibilă simțurilor noastre, Hume spune că noi „proiectăm" cauzalitatea în exterior, în lume, pe baza percepției repetate a lucrurilor, care are loc în succesiuni temporale apropiate. Nu observăm direct „cauzalitatea" în afară – și nici nu putem face asta. Da, în afară se întâmplă multe, dar ceea ce trăim noi drept cauzalitate este, de fapt, inferență perceptivă, în același mod în care toate percepțiile noastre sunt proiecții ale așteptărilor structurate din creierul nostru către și în mediul nostru senzorial – exerciții din categoria contribuției privitorului. După Hume, mintea are înclinația de a se împrăștia spre și în lume, în așa fel încât noi „marcăm și înfrumusețăm" obiectele naturale „cu culorile împrumutate din sentimentul nostru interior". Și nu e vorba doar despre culori: forme, mirosuri, obiectualitate, schimbări, durate și cauzalitate – caracteristici din prim plan, dar și din fundal ale lumii perceptive – toate sunt proiecții humeene, aspecte ale unei halucinații controlate.

De ce ne trăim construcțiile perceptive ca fiind obiectiv reale? Din perspectiva halucinației controlate, scopul percepției este de a ghida acțiunea și comportamentul – de a promova șansele organismului de a supraviețui. Percepem lumea nu *așa cum este*, ci *cum ne este nouă util*. Astfel, are sens faptul că proprietățile fenomenologice – calitatea de

Scanned with OKEN Scanner

roșu, calitatea de scaun, de Cilla Black, de cauzalitate - *par a fi* obiective, veridice, proprietăți ale unui mediu extern existent. Putem răspunde mai repede și mai eficient la ceva ce se întâmplă undeva dacă percepem acel ceva ca fiind real. Acolo-afară-ul inerent din experiența noastră perceptivă a lumii este, cred, o caracteristică necesară a modelului generativ care e capabil să-și anticipeze fluxul aferent, senzorial, pentru a putea ghida corect comportamentul.

Cu alte cuvinte, chiar dacă proprietățile perceptive depind de modelele generative top-down, noi nu trăim modelele *ca pe* modele. Mai degrabă percepem *cu* și *prin* modelele noastre generative și, făcând astfel, dintr-un simplu mecanism se naște o lume structurată.

La începutul acestei cărți promiteam că, urmărind abordarea problemei reale, vom ajunge la problema grea a de ce și cum orice fel de mecanism fizic ar trebui să genereze, să corespundă sau să fie totuna cu experiența conștientă. Facem vreun progres?

Da, facem. Pornind de la principiul conform căruia creierul trebuie să infere cauzele ascunse ale input-urilor sale senzoriale, am ajuns la o nouă înțelegere a de ce și cum anume universul nostru interior este populat cu orice, de la căni de cafea la culori și proprietăți de cauzalitate – care *par să fie* proprietăți ale unei realități obiective exterioare, unde acest *pare-să-fie* este chiar o proprietate a inferenței perceptive. Și chiar proprietatea de „a părea să fie real" aduce argumente suplimentare convingerilor interioare dualiste despre cum anume experiența conștientă și lumea fizică se înrudesc, convingeri care, în schimb, conduc la ideea problemei grele. Asta deoarece percepțiile noastre au trăsătura fenomenologică de „a fi reale", și este extrem de dificil de apreciat că, de fapt, experiențele perceptive nu corespund neapărat – sau vreodată – direct cu lucruri care există independent de minte. Un scaun are o existență independentă de minte; *calitatea de scaun*, însă, nu.

Scanned with OKEN Scanner

Odată ce realizăm asta, putem recunoaște despre problema grea că e mai puțin grea sau, poate, chiar o non-problemă. Privind și altfel lucrurile, problema grea a conștienței pare grea mai ales dacă interpretăm conținutul experienței noastre perceptive ca existând cu adevărat acolo, afară. Adică exact ce ne încurajează automat să facem fenomenologia percepției conștiente normale.

Așa cum s-a întâmplat cu studiul vieții acum un secol, nevoia de a găsi un „sos special" pentru conștiență scade direct proporțional cu îmbunătățirea capacității noastre de a distinge diferite aspecte ale experienței conștiente și de a le descrie în termenii mecanismelor fiecăreia. Dizolvarea problemei grele diferă de rezolvarea sau de respingerea ei, dar e cel mai bun mod de a progresa, cu mult mai bun decât venerarea conștienței ca pe un mister magic sau decât a o trata ca pe o non-problemă metafizică iluzorie. Iar misiunea noastră capătă substanță atunci când considerăm că nu numai experiențele de trăire a lumii reprezintă constructe perceptive.

E timpul să întrebăm cine sau ce anume se ocupă de toată această percepție.

Scanned with OKEN Scanner

III

SINE

Scanned with OKEN Scanner

7
Delir

În vara lui 2014, mama mea a intrat în stare vegetativă pe când se afla la unitatea urgențe de la secție chirurgie a spitalului John Radcliffe din Oxford. Avea o encefalopatie nediagnosticată – o boală la creier. Cauza nu a fost niciodată pe deplin elucidată; fusese internată pentru cancer la intestine, iar problemele neurologice au fost neașteptate. M-am grăbit să mă întorc de la o conferință la care fusesem invitat, în Brisbane, temându-mă de ce e mai rău. Și-a revenit, încet, dar amintirea acelei disoluții la care am asistat, cu fiecare oră, mi-a rămas întipărită în minte. Ea nu-și amintește, din fericire, mai nimic.

Patru ani mai târziu, în vara lui 2018, ne aflam în mijlocul unui val neobișnuit de căldură și al campionatului Cupei Mondiale. De data asta, fără starea vegetativă. În schimb, mama mea – de 83 de ani la acea vreme – suferea de ceea ce aveam să descopăr că se numește „delir indus-de-spital", un alt tip de disoluție a sinelui și a lumii din jurul ei. Cu două săptămâni mai devreme, fusese adusă de urgență la spitalul Radcliffe pentru o recurență a durerilor intestinale severe. Două zile după internare, în timp ce aștepta să vadă dacă problema la intestine se rezolvă fără chirurgie, a început să aibă halucinații și închipuiri. Am condus din Brighton să-i fiu alături.

Delir – cuvânt care vine din adâncurile secolului al XVI-lea – provine din latinescul *delirare*, a devia, a fi deranjat. Dicționarul îl explică drept „o stare confuzională acută, caracterizată de neliniște, iluzii și incoerență": spre deosebire de demență, care e cronică și degenerativă,

Scanned with OKEN Scanner

delirul este, de obicei, trecător. Crește și descrește, uneori pe parcursul mai multor săptămâni. În capul meu, cuvântul evocă azilurile victoriene, așa că am fost surprins când l-am auzit ca diagnostic de secol XXI într-un spital din Anglia. Reflectând, însă, dincolo de surpriză, e clar că psihiatria mai are drum lung de parcurs.

În ce o privește pe mama, definiția din dicționar a fost corectă. Am găsit-o pe secție, cocoșată în scaun, fără zâmbet, despletită, cu ochii goi. Mi-a povestit despre oameni care se târăsc pe pereți, că nu-și amintește unde e și nici de ce e acolo. Nu mai avea simțul realității.

Apogeul avea să fie atins vineri. Nu mai avea încredere în nimeni și era convinsă că e victima unui experiment de proporții și plin de cruzime și că noi – inclusiv eu, mai ales pentru că, în paranoia ei, eram adesea stăpânul inelelor – îi inducem deliberat aceste halucinații prin medicamentele pe care insistam să le ia, din motive obscure și malefice. Fermecătoare și bună în viața obișnuită, ajunsese să se răstească furioasă la asistente, cerând să părăsească spitalul, încercând de mai multe ori să scape și ordonând medicilor să-l alunge pe fiul ei, cercetătorul nebun. Asta nu e mama mea. E o femeie care arată ca mama, dar nu e ea.

Factorii de risc pentru delirul indus-de-spital includ epilepsia, infecția, chirurgia majoră (sau o situație care necesită chirurgie majoră), febra, deshidratarea, privarea de hrană și de somn, efectele secundare ale medicației și – important – lipsa de familiaritate a locului. Toate acestea s-au întâmplat mamei mele. Lipsa de familiaritate a locului este motivul pentru care acest tip de delir se numește „indus-de-spital".

Puține lucruri pot provoca o disociere de lumea reală așa cum se întâmplă la urgențe, în secțiile de chirurgie. Semnalele acustice ritmice și luminile pâlpâind continuu, lipsa aproape cu desăvârșire a legăturii cu lumea exterioară – o privire pe geam, dacă ai noroc – o întreagă lume

Scanned with OKEN Scanner

restrânsă la un pat, un scaun și, poate, un coridor. Alții ca tine trecând pe lângă, în diferite stadii de suferință, echipe de asistente, asemănătoare dar diferite, doctori debutanți, doctori care consultă. În fiecare zi la fel. Delirul, deși urgență medicală, nu este recunoscut sau tratat ca atare. Pacienții sunt internați cu o anumită boală, *aceasta* fiind ținta tratamentului, nu orice problemă a minții sau a creierului care se poate dezvolta pe parcurs.

Aproape o treime dintre pacienții în vârstă internați cu boli acute ajung să dezvolte delir indus-de-spital, iar procentul e mai mare în cazul celor care trec prin operații. Chiar dacă se ameliorează în timp, consecințele pe termen lung pot fi severe, incluzând reducerea capacității cognitive, risc de moarte în lunile următoare și risc crescut de a dezvolta episoade ulterioare de delir și demență. Am vizitat casa de la țară a mamei, aceeași casă în care am crescut, ca să-i aduc câteva obiecte familiare, în speranța că o vor ajuta să se reorienteze, să se agațe de ceva cunoscut. O fotografie înrămată, ochelarii, un pulover, un leu de pluș de pe vremea când eram copil.

Închipuirile sunt rareori întâmplătoare. Cele care caracterizează delirul mamei mele au o logică răstălmăcită. Ea crede – *știe* – că e victima unui experiment diabolic, care-i implică pe toți cei din jurul ei și la care eu sunt complice. Da, fac experimente pe oameni și, când sunt în spital, devin un soi de identitate fluidă, parte fiu, parte medic, încercând să o alin în timp ce mă uit peste consemnările medicale și discutând cu doctorii despre encefalopatii paraneoplastice și alte lucruri oribile. Creierul încearcă mereu să găsească o rezolvare, să-și formuleze cea mai bună estimare.

Câteva dintre alterările comportamentului ei sunt chiar subtile. Propozițiile ei sunt din cuvinte rostite separat, nu curg ca un flux coerent. „Nu-mi găsesc ochelarii, nu știu unde sunt". Această stare se menține multe zile după ce delirul primar s-a atenuat. Apare și dispare și ea. În seara

Scanned with OKEN Scanner

asta a recăzut, altă regresie, spulberându-mi speranțele de a o vedea acasă curând.

Propria mea viață începe să pară ireală pe la margini. Sunt singura familie apropiată pe care o are, deci mie îmi revine sarcina de a fi acolo pentru ea. Dimineți și seri petrecute la spital, după-amiezele, dacă am noroc, încercând să ajung la zi cu munca, plimbându-mă și înotând în Tamisa. În astfel de după-amiezi mă îndrept spre Port Meadow, o pajiște mare, plină de gâște, lebede, vaci și cai sălbatici. Un loc improbabil, chiar și printre lumile stratificate ale Oxfordului. Valul de căldură ține de săptămâni, așa că acele câmpii, de obicei noroioase, seamănă acum cu savana africană. Azi dimineață m-a alergat un cal sălbatic iar inima mi-a luat-o razna pe când traversam podul de peste calea ferată către oraș.

Este ziua a paisprezecea în spital, pentru mama mea, și ziua a douăsprezecea pentru mine. Confuzia ei acută a trecut, deși e încă schimbată și starea ei fluctuează. Acum e uluită când îi spun că credea că fac experimente pe ea, că încerca să mă înlăture. Îi țin mâna, îi spun că totul va fi în regulă și sper că se va întoarce pe deplin la ea însăși.

Dar ce *este* un „sine"? Este acel ceva de la care se poate pleca și la care să te întorci?

Și sinele este altceva decât pare să fie.

Scanned with OKEN Scanner

8
Așteaptă-te pe tine

Poate să pară că sinele – sinele tău – este acel „ceva“ care percepe. Numai că lucrurile nu stau așa. Sinele este o altă percepție, o altă halucinație controlată, de un anume fel, însă. De la simțul identității personale – cum e acela de a fi om de știință sau fiu – la experiența de a avea un corp și de *a fi*, pur și simplu, un corp, multele și variatele elemente ale identității personale sunt, de fapt, cele mai bune estimări bayesiene, proiectate de evoluție ca să te mențină în viață.

Să ne începem explorarea sinelui cu o scurtă excursie în viitor. Peste un secol, dispozitivele teleportoare se vor fi inventat, creând replici perfecte ale oricărei ființe umane. La fel ca mașinile din Star Trek, scanează o persoană în cel mai mic detaliu – până la aranjamentul fiecărei molecule – și, cu ajutorul informației din scanare, construiește o a doua versiune a persoanei respective în altă parte, de exemplu, pe Marte.

După temerile inițiale, oamenii s-ar obișnui rapid cu această tehnologie ca mijloc eficient de transport. S-ar obișnui și cu faptul că, odată creată replica, originalul este vaporizat imediat – procedură absolut necesară, fără de care ar apărea o explozie de persoane identice. Din punctul de vedere al unei călătoare, pe care o numim Eva, nu există nicio problemă practică. După câteva asigurări din partea operatorului, Eva simte doar că a dispărut din locul X (Londra), reapărând în Y (Marte) într-o secundă.

Într-o zi, apare o defecțiune. Modulul de vaporizare din Londra nu funcționează cum trebuie, iar Eva – care e în

Scanned with OKEN Scanner

Londra, oricum – simte că nu s-a întâmplat nimic și că încă se află în dispozitivul de transport. O defecțiune minoră. Mașina trebuie să facă reboot și să reia operațiunea, sau să o lase, poate, pe a doua zi. Însă, un tehnician se repede în încăpere, cu o armă. Mormăie ceva de genul „nu te îngrijora, ai fost transportată în siguranță pe Marte, așa cum e normal, reglementările spun, însă, că trebuie să... și, uite, ai semnat aici de consimțământ...“ El ridică încet arma și Eva are un sentiment pe care nu l-a mai avut niciodată, că, poate, tâmpenia asta de teleportare nu e chiar atât de sigură, până la urmă.

Scopul acestui experiment de gândire, numit „paradoxul teleportării“, este de a dezvălui câteva dintre prejudecățile pe care majoritatea le avem despre ce anume înseamnă *a fi un sine*.

Paradoxul teleportării a dus la formularea a două probleme filosofice. Prima, despre conștiența-în-general, se referă la cât de siguri suntem că replica va avea experiențe conștiente, sau dacă va fi vorba de un echivalent perfect funcțional, dar fără univers interior. Nu mi se pare foarte interesantă această problemă. Dacă replica este recreată cu suficiente detalii – cu fiecare moleculă identică – atunci nu există niciun motiv să mă îndoiesc că va fi și conștientă, în exact același mod în care era originalul. Dacă replica nu este perfect identică, atunci ne întoarcem la argumentele despre diferitele tipuri de zombi filosofici – nemaifiind nevoie să reluăm totul.

Problema mai interesantă este cea a identității personale. Este Eva de pe Marte (să-i zicem Eva 2) *aceeași persoană* ca Eva 1 (Eva care e încă în Londra)? Tentația e să spui da, este: Eva 2 va simți în toate felurile exact cum Eva 1 s-ar fi simțit dacă chiar ar fi fost transportată din Londra pe Marte. Ceea ce pare să conteze pentru acest tip de identitate personală este continuitatea psihologică, nu cea

Scanned with OKEN Scanner

fizică.* Dar dacă Eva 1 nu a fost vaporizată, *care dintre ele este adevărata Eva*?

Cred că răspunsul corect – dar și ciudat – este că *ambele* sunt adevărata Eva.

Considerăm, instinctiv, trăirile sinelui ca fiind diferite de ale lumii. Când vine vorba de experiența de *a fi tu*, pare greu să reziști ideii că exprimă, mai degrabă, o proprietate autentică a felului în care sunt lucrurile – în acest caz, *un sine real* – decât o serie de percepții. O consecință instinctivă a presupunerii existenței unui sine real este convingerea că nu poate fi decât un singur sine, nu doi, trei sau mai mulți.

Ideea că sinele ar fi cumva indivizibil, imuabil, transcendental, *sui generis* face parte din idealul cartezian al sufletului imaterial și are, încă, o rezonanță psihologică profundă, mai ales în societățile vestice. Dar a fost tratată cu scepticism de către filosofi și de practicanți ai diverselor religii, iar mai recent, și de psihonauții psihedelici, medici și specialiști în neuroștiință.

În a sa *Critică a rațiunii pure*, Kant consideră eronat conceptul de sine ca „substanță simplă", iar Hume consideră sinele ca fiind o „sumă" de percepții. Mult mai recent, filosoful german Thomas Metzinger a scris o carte extraordinară, *Being No One* – o deconstrucție convingătoare a sinelui singular. Budiștii susțin de multă vreme că un sine permanent nu există, ei străduindu-se să ajungă la o stare de conștiență complet lipsită de sine. Ceremoniile Ayahuasca din America de Sud, din ce în ce mai frecvente și în alte părți, reușesc să estompeze simțul sinelui până la

* Chiar și fără teleportare, celulele corpului nostru se schimbă permanent, fiind înlocuite cu totul la fiecare circa 10 ani – o corabie biologică a lui Tezeu. Asta nu pare să aibă impact asupra simțului identității noastre personale.

Scanned with OKEN Scanner

dispariție, cu ajutorul unor amestecuri impresionante de ritual și dimetil-triptamină.

În neurologie, Oliver Sacks și alții au consemnat nenumăratele feluri în care sinele dispare ca urmare a unor boli ale creierului sau a lezării lui, în timp ce pacienții cu creierul divizat – pe care i-am întâlnit în capitolul 3 – sugerează că un sine poate deveni doi. Cel mai curios caz dintre toate este cel al gemenilor siamezi, legați și la nivel cerebral. Ce înseamnă, deci, un sine individual, atunci când unul dintre gemeni poate simți când celălalt bea un suc de portocale?

A fi tu nu e chiar așa de simplu cum pare.

Întorcându-ne la dispozitivul de teleportat, Eva 1 reușește să scape de tentativa de omor a tehnicianului și se adaptează noii ei situații, în timp ce Eva 2 nu are nici cea mai mică idee de întreaga dramă care se desfășoară pe Terra.

Chiar dacă ambele Eva erau identice obiectiv și subiectiv în momentul replicării, identitățile lor încep să diveargă. La fel ca gemenii identici care își parcurg, fiecare, propriile trasee în viață, procesul se complică în timp. Chiar dacă Eva 1 ar sta lângă Eva 2, input-urile senzoriale ar fi ușor diferite, ceea ce ar duce la diferențe subtile de comportament și, fără să-ți dai seama, Eva 1 și Eva 2 trăiesc lucruri diferite, depozitând amintiri diferite, devenind persoane diferite.

Aceste complexități ale identității personale ne caracterizează și pe noi, în diferite moduri. Identitatea mamei mele s-a schimbat dramatic în timpul delirului și, chiar dacă și-a revenit, cel puțin pentru mine, pare și diferită de, dar și aceeași cu cea care era înainte – exact ca în cazul celor două Eve. Relația dintre Eva 1 și Eva 2 poate fi oarecum ca relația dintre tine acum și tine de acum zece ani – sau de peste zece ani.

Scanned with OKEN Scanner

Când vine vorba de cine ești sau de cine sunt – acel eu care este subiectiv și obiectiv „Anil Seth" – lucrurile nu sunt atât de simple cum par la prima vedere. Dintr-un punct de vedere, simțul identității personale – acel „eu din spatele ochilor"/I behind the eyes – este doar un aspect al felului în care „a fi un sine" apare în conștiență.

Iată cum îmi place să desfac, până la ultimul, elementele sinelui uman.

Există experiențe ale *sinelui întrupat/embodied selfhood* care se leagă direct de corp. Acestea includ impresia de identificare cu un anume obiect, care se întâmplă să fie corpul tău – avem un simț al proprietății asupra corpului nostru și care nu se aplică altor obiecte. Emoțiile și stările de spirit sunt și ele aspecte ale sinelui întrupat, la fel și stările de excitare și alertă. Coborând frecvența, ajungem la trăiri mai profunde, fără formă definită, de a fi pur și simplu un organism viu întrupat – de *a fi un corp* – fără o extindere spațială clar definită sau context specific. Vom reveni ulterior la acest nivel de bază al sinelui. Deocamdată, e suficient să îl gândim ca fiind „sentimentul de a fi viu".

Deplasându-ne dincolo de corp, există experiența de a percepe lumea dintr-un anumit punct de vedere, de a avea o perspectivă la persoana I – un punct subiectiv de la care se formează experiența perceptuală și care pare să se afle, de obicei, în cap, localizat cumva între ochi și în spatele frunții. Acest *sine-perspectivă* este ilustrat cel mai bine în autoportretul fizicianului austriac Ernst Mach, lucrare cunoscută și ca „Vedere din ochiul stâng".

Experiențe precum *voința*, intenția de a face lucruri – intenția – și de a fi cauza lucrurilor care se întâmplă – agency – sunt și ele esențiale pentru sine. Acesta este *sinele volițional*. Când ne referim la „liber arbitru", despre aceste aspecte ale sinelui vorbim de fapt. Pentru mulți, noțiunea

Scanned with OKEN Scanner

de „liber arbitru“ surprinde acel aspect de a-fi-un-sine la care nu vor să renunțe în fața științei.

Fig. 16: Ernst Mach, Autoportret (1886)

Toate aceste modalități de a-fi-un-sine pot fi la locul lor înainte de orice concept de identitate personală – identitate care poate fi asociată cu un nume, un trecut și un viitor. Așa cum am văzut în paradoxul teleportării, pentru ca identitatea personală să existe, e nevoie de un trecut personalizat, un fir de amintiri autobiografice și un viitor proiectat.

Atunci când apare, acest simț al identității personale poate fi numit *sine narativ*. Odată cu apariția lui, dobândim și capacitatea de a trece prin emoții sofisticate precum regretul, total diferit de simpla dezamăgire. (Noi, oamenii, putem suferi și de „regret anticipativ“ – siguranța că voi face un rău și că, deși știu asta, îl fac oricum, iar eu și ceilalți vom suferi consecințele.) Aici putem vedea cum se ramifică și interacționează diferite niveluri ale sinelui – apariția

Scanned with OKEN Scanner

identității personale se schimbă, dar este și parțial definită de un domeniu tot mai vast de stări emoționale oferite.

Sinele social se referă la modul în care percep cum mă percep ceilalți. Este acea parte din mine care apare ca urmare a încastrării ființei mele în rețeaua socială. Sinele social apare treptat în timpul copilăriei și continuă să evolueze de-a lungul vieții, dezvoltându-se diferit în situații precum autismul. Sinele social aduce cu el întreaga gamă de posibilități emoționale, de la noi modalități de a te simți rău – precum vina și rușinea –, la modalități de a te simți bine – precum mândria, iubirea, apartenența.

Pentru fiecare dintre noi – în circumstanțe normale – diversele elemente ale sinelui sunt legate împreună, într-o singură piesă, toate subsumate într-o experiență cuprinzătoare, unificată – aceea de *a fi tu*. Caracterul unificat al acestei experiențe poate părea atât de natural – la fel de natural precum modul în care se leagă culoarea și forma atunci când privești un scaun roșu – încât e ușor să-l consideri implicit.

Ar fi o greșeală să consideri astfel. Așa cum trăirile roșului nu sunt indicatori ai existenței unui „roșu" exterior, experiențele sinelui unificat nu dovedesc existența unui „sine real". Într-adevăr, ideea de sine unificat se poate demonta foarte ușor. Simțul identității personale, construit pe sinele narativ, se poate eroda până la dispariție completă în cazul demenței sau al amneziei severe, fiind distorsionat și deformat în cazul delirului, indus sau nu de spital. Sinele volițional o poate lua razna în situații precum schizofrenia și sindromul mâinii străine, când conexiunea cu propriile acțiuni scade, sau în mutismul akinetic, o tulburare în care omul nu mai comunică cu împrejurimile în niciun fel. Experiențele în afara corpului și alte tulburări de disociere afectează sinele-perspectivă, în timp ce tulburările din gama proprietății asupra corpului, de la sindromul membrului fantomă – senzația persistentă, uneori dureroasă, percepută

Scanned with OKEN Scanner

de la un membru care nu mai există – la somatoparafrenie – senzația că unul din membrele tale aparține altcuiva. În xenomelie – o formă extremă a somatoparafreniei – individul simte o dorință nestăvilită de a-și amputa un braț sau un picior, remediu drastic pe care, în rare ocazii, îl duc până la capăt.

Sinele nu este o entitate imuabilă care bântuie în spatele ochilor-ferestre, privind în afară și controlând corpul precum un pilot, avionul. Experiența de *a fi eu* sau de *a fi tu* este o percepție în sine – sau, mai degrabă, o sumă de percepții –, un bagaj strâns împachetat de predicții codificate neuronal, destinat supraviețuirii tale. Și asta este, cred, tot ceea ce avem nevoie să fim – noi înșine.

Fie experiența de a te identifica cu un anume obiect din afară, care e corpul tău. Natura schimbătoare și precar asamblată a acestor experiențe se manifestă nu numai în tulburări precum somatoparafrenia și sindromul membrului fantomă, ci poate fi evidențiată și cu ajutorul unor experimente simple de laborator. Cel mai cunoscut este „iluzia mâinii de cauciuc", descris pentru prima dată în urmă cu mai bine de 20 de ani, acum o piatră de temelie a cercetării pe tema întrupării/embodiment.

Iluzia mâinii de cauciuc se poate verifica cu ușurință – avem nevoie de un voluntar, câteva bucăți de carton, pentru a forma un mic perete separator, două pensule și o mână de cauciuc. Aranjamentul se poate vedea în ilustrația din Fig. 17. Voluntarul își așază mâna (adevărată) de o parte a separatorului de carton, în afara câmpului său vizual. Mâna de cauciuc e așezată în fața voluntarului, în același loc și în același fel în care ar fi așezată mâna reală. Experimentatorul mângâie cu o pensulă, de mai multe ori, ambele mâini – reală și de cauciuc. Atunci când mâinile sunt pensulate simultan,

Scanned with OKEN Scanner

voluntarul are iluzia că mâna de cauciuc îi aparține, chiar dacă nu e așa. Atunci când pensularea nu e simultană, iluzia nu se formează, iar voluntarul nu va asimila mâna de cauciuc în experiența trăirii propriului corp.

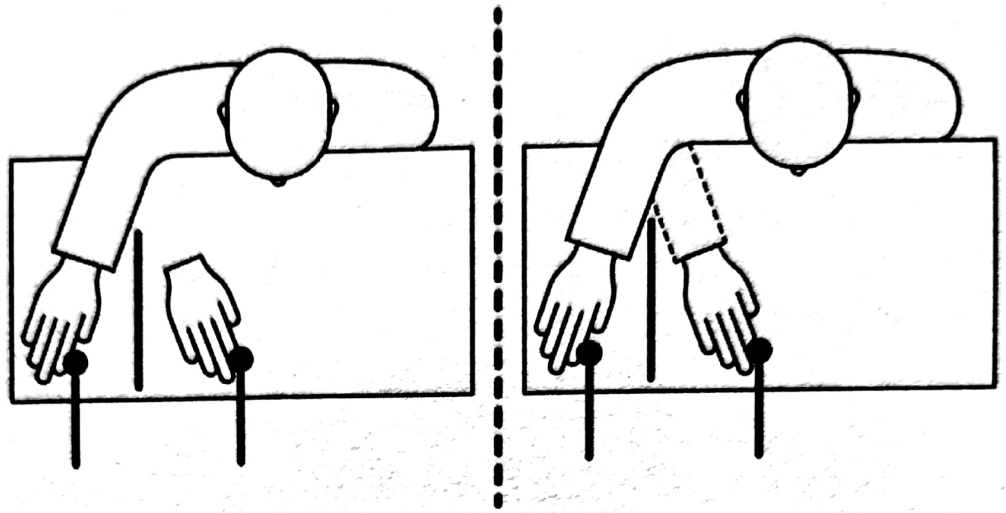

Fig. 17: Iluzia mâinii de cauciuc. Atunci când mâna de cauciuc și mâna reală sunt pensulate simultan (stânga), senzația de posesie a corpului poate devia, astfel încât mâna de cauciuc să ajungă să fie considerată parte a corpului (dreapta).

Pentru unii, aceasta este o descriere potrivită pentru ceea ce se întâmplă, iar senzația că o mână evident falsă ar putea fi considerată cumva-dar-nu-pe-de-a-ntregul parte a propriului corp este, fără îndoială, ciudată. Acestea fiind spuse, experiența în sine variază de la persoană la persoană. O cale de a investiga asta este de a amenința brusc mâna de cauciuc cu un ciocan sau cu un cuțit – reacția va fi puternică dacă ești sub influența iluziei.

Iluzia mâinii de cauciuc se potrivește foarte bine cu ideea că simțul proprietății asupra corpului tău reprezintă un anume tip de halucinație controlată.

Ideea e că, în cazul pensulării simultane, combinația dintre *a vedea* mâna de cauciuc și *a simți* simultan (fără a vedea) atingerea pe mâna reală oferă suficiente dovezi senzoriale pentru ca creierul să formuleze cea mai bună estimare perceptivă conform căreia mâna din cauciuc este, cumva, parte a corpului. Asta se întâmplă atunci când

Scanned with OKEN Scanner

pensularea se face sincron – nu și atunci când e asincronă – ca urmare a unei așteptări precedente conform căreia semnalele senzoriale care ajung simultan trebuie că au o sursă comună – mâna de cauciuc.

Nu numai părți ale corpului pot fi simțite diferit. Se poate întâmpla și cu întregul corp – originea perspectivei la persoana I.

În 2007, în revista Science apăreau, aproape în același timp, două articole. Ambele descriau cum anume puteau fi folosite noi metode din realitatea virtuală pentru a genera experiențe asemănătoare celor „extracorporale". Experimentele se bazau pe iluzia mâinii de cauciuc, dar au fost extinse la nivelul întregului corp. Într-unul din studii, realizat de un grup din Lausanne, condus de Olaf Blanke, voluntarii purtau pe cap un dispozitiv cu display prin care puteau vedea reprezentarea în realitate virtuală a spatelui propriului corp de la o distanță de circa doi metri (vezi imaginea de mai jos). Din această perspectivă, și-au văzut corpul virtual pensulat fie sincron, fie asincron cu propriul

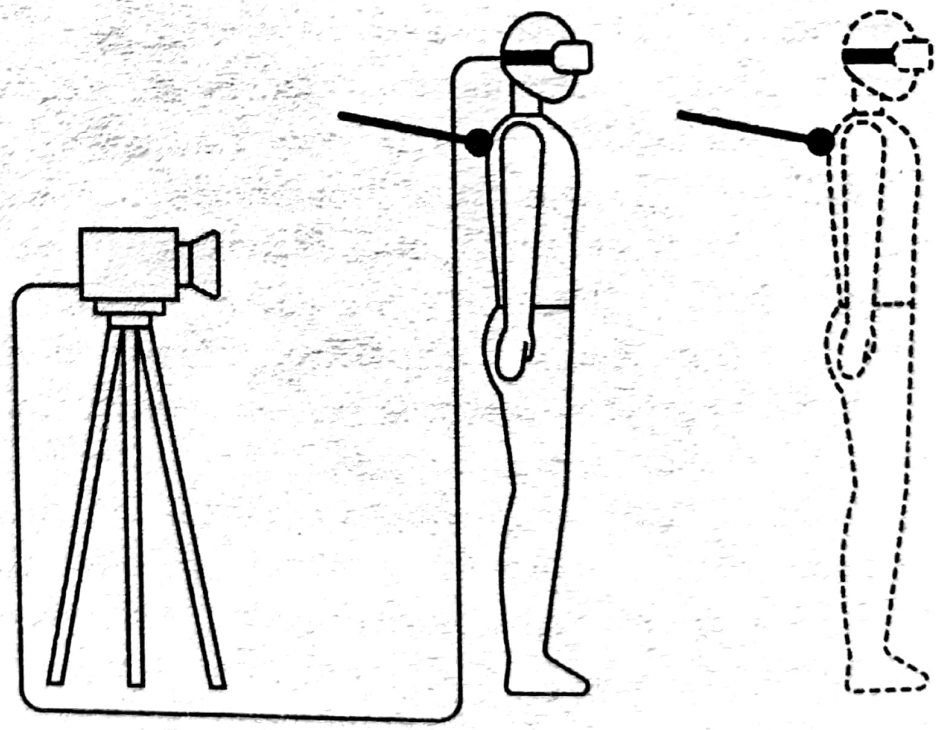

Fig. 18: Crearea „iluziei întregului corp"

Scanned with OKEN Scanner

corp, cel real. Atunci când pensularea era sincronă, cei mai mulți participanți au raportat că percep corpul virtual ca și cum ar fi propriul „lor" corp și, când li s-a spus să meargă acolo unde simțeau că li se află corpul, s-au deplasat în locul în care se afla corpul virtual.

În același mod în care iluzia mâinii de cauciuc indică o flexibilitate, adaptată în fiecare moment, a simțului proprietății asupra corpului, experimente ca acestea – numite „iluzii ale întregului corp" – sugerează că proprietatea subiectivă asupra întregului corp, precum și localizarea perspectivei la persoana I, pot fi manipulate din mers. Aceste experimente vin cu dovezi fascinante despre faptul că trăirile de tip „ce este corpul meu" pot fi disociate, până la un punct, de trăirea de tip „unde mă aflu".

Ideea conform căreia propria perspectivă la persoana I poate părăsi corpul fizic sub forma experiențelor de decorporare – OBE (out-of-body experiences) – este înscrisă profund în istorie și cultură. Raportări de OBE sau de experiențe similare OBE în timpul experiențelor traumatice, vecine cu moartea, în sălile de operații și la limita crizelor epileptice au alimentat convingerile despre o esență-a-sinelui imaterială. Până la urmă, dacă te poți privi din afară, atunci baza conștienței trebuie că e separabilă de creierul tău?

Numai că nu este nevoie să derapăm în astfel de abordări dualiste dacă vom considera, despre perspectiva la persoana I, că este o altă specie de inferență perceptuală. Această abordare este susținută nu numai de experimentele de realitate virtuală ale unor Olaf Blanke și alții, ci și de studii de stimulare cerebrală, dintre care primele cu adevărat semnificative au fost făcute în anii 40, de neurologul canadian Wilder Penfield.

Printre pacienții lui Penfield era o femeie, G.A., care, atunci când i se stimula electric girusul temporal superior drept – parte a lobului temporal – exclama brusc „am o

Scanned with OKEN Scanner

senzație ciudată că nu mă aflu aici... ca și cum aș fi jumate aici, jumate în altă parte". Blanke însuși a devenit fascinat de OBE atunci când un pacient de-al lui, stimulat electric într-o parte a creierului asemănătoare – girusul angular, la joncțiunea dintre lobii temporal și parietal – a raportat o experiență similară: „mă văd întins în pat, de sus, dar doar picioarele sunt cu adevărat vizibile".

Factorul comun în astfel de cazuri îl reprezintă activitatea neobișnuită din regiunile cerebrale care procesează input-urile vestibulare (sistemul vestibular se ocupă de simțul echilibrului) și care sunt implicate și în integrarea multi-senzorială. Se pare că atunci când activitatea normală din aceste sisteme este perturbată, creierul poate ajunge la o „cea-mai-bună-estimare" neobișnuită despre localizarea perspectivei sale la persoana I, chiar și atunci când alte aspecte ale sinelui rămân nealterate.

Experiențele similare OBE care însoțesc, uneori, crizele epileptice, pot fi urmărite și ele până la perturbările care au survenit în aceste procese. Experiențele respective sunt împărțite, de obicei, în *halucinații autoscopice* – în care vezi împrejurimile dintr-o perspectivă diferită – și *halucinații heautoscopice* (denumite și halucinații doppelgänger/dublură) – în care te vezi pe tine dintr-o altă perspectivă. Documentația extensivă despre experiențe de acest tip – care se întinde pe ultimele sute de ani – reprezintă dovezi suplimentare ale maleabilității acestei perspective la persoana I.*

Orice descriere a unor experiențe aparent supranaturale sau bizare, cum sunt OBE, trebuie luată în serios. Cei care le raportează chiar le trăiesc. Omul are experiențe extracorporale de mii de ani, numai că asta nu se traduce prin existența unor suflete imateriale care părăsesc corpul fizic. Descrierile unor astfel de experiențe dezvăluie

* Halucinațiile doppelgänger au fost popularizate de Fiodor Dostoievski în nuvela sa, *Dublura*. Se știe că Dostoievski suferea de epilepsie severă.

Scanned with OKEN Scanner

asocierea perspectivelor la persoana I în maniere mult mai complexe, temporare și precare decât cele la care vom avea vreodată acces subiectiv direct.

În lumea virtuală, capacitatea de a modifica perspectiva la persoana I generează câteva aplicații fascinante, multe dintre ele având la origine iluzia numită, contrariant, „schimb de corpuri", descrisă în 2008 într-un studiu condus de cercetătorul suedez Henrik Ehrsson. În aranjamentul experimental pentru schimbul de corpuri, doi subiecți au montate pe cap echipamente cu display, dotate cu câte o cameră video. Schimbând input-urile către camere între cele două echipamente, fiecare subiect se poate vedea din perspectiva celuilalt. Efectul se manifestă cu adevărat în momentul în care sunt puși să-și strângă mâinile. Vederea și strângerea simultană a mâinii înseamnă stimulare multisenzorială astfel că, la combinarea cu așteptările top-down, fiecare persoană simte că se află, oarecum, în corpul celeilalte, strângându-și propria mână. Această experiență te pune – e adevărat, virtual – în papucii celuilalt.

Am încercat și eu un schimb virtual de corpuri la o mică adunare în Ojai, California, în iarna lui 2018. Eram acolo împreună cu Daanish Masood, negociator de pace de la Națiunile Unite, dar și cercetător pe realitate virtuală. Masood lucrase îndeaproape, câțiva ani, cu BeAnotherLab, creația lui Mel Slater, cercetător în neuroștiință stabilit în Barcelona. Scopul BeAnotherLab este acela de a adapta tehnologia de schimb de corpuri într-o generație nouă de dispozitive pentru „generarea empatiei". Ideea e că, probând percepția lumii din corpul virtual al celuilalt, s-ar putea genera empatie pentru celălalt.

Daanish și-a adus echipa la Ojai pentru a face demonstrații cu sistemul lor, numit Mașina de a fi Altul.

Scanned with OKEN Scanner

Pe lângă principiul de bază al schimbului între corpuri, instalația are și o coregrafie inteligentă, care amplifică efectul. Doi participanți își pun casca, apoi privesc mai întâi în jos, în poală, pentru a vedea corpul celuilalt, în loc de al lor. Apoi fac o serie de mișcări coordonate, după instrucțiuni detaliate și, dacă le urmează cu grijă, noul lor corp va părea că răspunde comenzilor lor, accentuând experiența de a fi celălalt. După un timp, li se pun în față oglinzi, iar fiecare vede în oglinda din față imaginea celuilalt, ca și cum ar fi el însuși. În final, se dă la o parte draperia care separă cei doi participanți iar ei se uită unul la celălalt ca și cum ar fi în corpul celuilalt, înainte să se apropie unul de altul și să se îmbrățișeze.

Atunci când mi-a venit rândul, am schimbat perspectiva proprie cu cea a unei doamne bine, de peste 70 de ani. Experiența a fost neașteptat de convingătoare. Îmi amintesc cum priveam în jos, flexându-mi (i) mâna și observând – cu oarecare surpriză – încălțămintea sport strălucitoare pe care (ea) o purta(m). Oglinda și îmbrățișarea finală au fost impresionante – nu sunt sigur dacă din cauza experienței de a simți că populez corpul altcuiva, sau a experienței de a mă vedea din perspectiva altcuiva. Doar mai târziu, la cină, mi-am dat seama ce ciudat s-ar putea să fi fost pentru partenera mea să fie transportată brusc în perspectiva la persoana I a unui cercetător în neuroștiințe englez, cu origini mixte și cu încălțăminte anostă.

Mi se pare fascinant că aceste aspecte familiare și considerate implicite ale sinelui – proprietatea subiectivă asupra corpului și perspectiva la persoana I – pot fi manipulate atât de ușor, fie cu mâini de cauciuc, fie cu pensule sau cu tehnologii noi de realitate virtuală și augmentată. Există, totuși, limite până la care pot merge

Scanned with OKEN Scanner

aceste manipulări. Așa cum am menționat mai devreme, experiența tipică în iluzia mâinii de cauciuc este de a simți mâna falsă ca făcând parte, oarecum. din propriul corp, chiar dacă știi că nu e așa. Iar această experiență „tipică" variază considerabil de la om la om – mulți nu simt mare lucru. La fel și în cazul iluziilor întregului corp și a schimbului de corpuri.

Aceste manipulări experimentale ale simțului proprietății asupra corpului diferă, astfel, destul de mult, de iluziile vizuale clasice precum tabla de șah a lui Adelson din capitolul 4. În cazul tablei de șah, suntem atât de convinși perceptiv că pătratele au tonuri diferite de gri, încât ne surprinde – poate chiar uimește – când ni se dezvăluie faptul că au același ton de gri. Acest tip de surprindere, obișnuită în iluziile vizuale, nu se întâmplă aproape niciodată în cazul iluziei proprietății asupra corpului. Pentru mine, cea mai convingătoare iluzie corporală de până acum a fost cea de schimb de corpuri pe care am probat-o în Ojai – dar nu am ajuns nici pe departe să cred că sunt altcineva sau altundeva.

Slăbiciunea subiectivă a iluziilor de proprietate asupra corpului este subliniată de un studiu recent în care am fost implicat și care examinează rolul sugestibilității hipnotice în iluzia mâinii de cauciuc. Raționamentul din spatele acestui studiu – condus de psihologii Peter Lush și Zoltán Dienes – a fost acela că aranjamentul experimental al iluziei vine cu așteptări implicite puternice în legătură cu ce va fi experimentat, iar aceste așteptări sunt suficiente pentru a genera experiențe alterate de proprietate asupra corpului la unii subiecți. În sprijinul acestei idei am aflat că diferențele individuale în puterea iluziei corelează cu gradul de sugestibilitate al persoanei, atunci când sunt măsurate pe o scală standard de hipnotizabilitate. Cei care pot fi hipnotizați ușor au raportat senzație puternică de proprietate (în cazul pensulării sincrone), în timp ce

Scanned with OKEN Scanner

subiecții greu hipnotizabili aproape că nu simțeau nimic în acest sens.

Pe de o parte, aceste descoperiri se potrivesc frumos cu ideea că senzația de proprietate asupra corpului este o halucinație controlată, cât timp o sugestie hipnotică poate fi gândită ca o puternică așteptare top-down – chiar și una de care participantul nu e neapărat conștient. Pe de alta, reprezintă o provocare serioasă pentru cercetarea experimentală în acest domeniu, deoarece vine cu ideea că iluzia mâinii de cauciuc ar putea să fie rezultatul, cel puțin parțial, al sugestiei. Până când studiile iluziilor de corporalitate nu vor ține cont de sugestibilitate – ceea ce nu au făcut până acum – va fi complicat să afirmi orice despre mecanismele implicate. Se aplică fie că vorbim de mâini de cauciuc, de experiențe asemănătoare decorporalizării, de iluzia schimbării între corpuri sau orice altă situație în care subiecții se așteaptă – implicit sau explicit – la anumite experiențe în legătură cu corpul lor.

Există un contrast clar între aceste iluzii oarecum blânde ale proprietății asupra corpului și experiențele puternic alterate observate în condiții clinice, precum somatoparafrenia, xenomelia și sindromul membrului fantomă, sau trăirea experiențelor de decorporalizare asociate cu convulsiile, sau declanșate de stimularea directă a creierului. Aceste distorsiuni dramatice seamănă cu iluziile vizuale clasice prin aceea că sunt mult mai convingătoare atunci când vin din partea celor care au experiențe neobișnuite. Și din acest motiv oferă dovezi mult mai puternice că experiențele de corporalitate și perspectiva sunt, cu adevărat, construcții ale creierului.

Să trecem la subiectele identității personale și al apariției sinelui „narativ" și „social". Așa cum am văzut în cazul paradoxului teleportării, acestea sunt nivelurile la care o enti-

Scanned with OKEN Scanner

tate se trăiește pe sine ca fiind continuă de la un moment la următorul, de la o zi, sau săptămână, sau lună la următoarea și – într-o oarecare măsură – de-a lungul întregii ei vieți. Acestea sunt nivelurile sinelui la care are sens să asociem sinele cu un nume, cu amintiri din trecut și cu planuri de viitor. Acestea sunt nivelurile la care devenim conștienți că avem un sine – devenim cu adevărat *conștienți-de-sine.*

Aceste realizări ale sinelui sunt perfect disociabile de sinele întrupat. Multe animale non-umane, dar și bebelușii pot trăi întruparea sinelui sub forma identității personale. Și, dacă omul adult trăiește toate aceste forme ale sinelui într-o manieră integrată și unificată, atunci când aspectele narativ și social al sinelui sunt diminuate sau distruse, impactul poate fi devastator.

Clive Wearing este un muzicolog britanic renumit pentru editarea lucrărilor compozitorului renascentist Orlando di Lasso, pentru cariera de dirijor de cor în Londra și pentru regândirea conținutului muzical al BBC Radio 3 la începutul anilor 80. În martie 1985, la apogeul carierei, a suferit o infecție cerebrală devastatoare, o encefalită herpetică ce i-a distrus masiv hipocampul în ambele emisfere, consecința fiind una dintre cele mai profunde amnezii consemnate vreodată.*

Wearing avea probleme mari în a-și aminti lucruri din trecutul îndepărtat (amnezie retrogradă) și, mai ales, în fixarea de amintiri din trecutul imediat (amnezie anterogradă). Remarcabil, părea să existe într-un prezent perpetuu cuprins între 7 și 30 de secunde. Acum are peste 80 de ani și pare că-și trăiește viața sub forma unei serii continue de mini-treziri, ca și cum – la fiecare 20 de secunde sau aproximativ – tocmai își revine din comă sau din anestezie. Sinele lui narativ a fost anihilat.

* Hipocampul este o structură curbată, de mici dimensiuni, situată profund în lobul temporal medial, asociată de mult cu consolidarea memoriei. Denumirea provine din cuvântul în elină pentru „căluț de mare".

Scanned with OKEN Scanner

Tipul de memorie pe care a pierdut-o Clive este cea episodică, autobiografică – memoria evenimentelor localizate în timp și spațiu (episodică), inclusiv, cel mai important, acele evenimente care-l implică pe el (autobiografice). Jurnalele lui conțin texte îngrozitoare. Sunt pline cu descrieri repetate ale unei „prime" treziri, una după alta, cu aserțiuni anterioare – unele scrise cu doar câteva clipe înainte – tăiate și, uneori, șterse cu furie.

~~8:31 am Acum sunt cu adevărat, complet treaz~~
~~9:06 am Acum sunt perfect, copleșitor de treaz~~
9:34 am Acum sunt chiar treaz, la superlativ

Aceste jurnale, precum și conversațiile cu Clive, înregistrate de soția sa în cartea ei, *Forever Today* (Mereu azi), stau mărturie a afectării simțului identității lui personale ca urmare a lezării creierului. Incapacitatea lui de a construi o autonarațiune în timp înseamnă că ceea-ce-ar-trebui-să-fie-el a fost, pentru mai mult de 30 de ani, un continuu început de la zero, o prezență trecătoare fără un „eu" stabil prin preajmă, care să organizeze fluxul de percepții ale lumii și ale sinelui. Abandonat în prezent din cauza profunzimii amneziei lui, pierderea trecutului și a viitorului este atât de dizlocantă încât chiar și el s-a întrebat dacă e viu sau nu. Deborah Wearing scrie: „Clive avea mereu impresia că abia s-a trezit din inconștiență, pentru că nu avea nicio dovadă, în mintea lui, că ar fi fost treaz înainte... «Nu am auzit nimic, văzut nimic, atins nimic, mirosit nimic», spunea. «Parcă aș fi mort».„

În același timp, alte aspecte ale simțului sinelui au rămas neatinse la Clive. Nu are nicio problemă cu senzația de proprietate asupra corpului, cu originea perspectivei sale la persoana I, nici cu execuția de acțiuni voluntare. Iubirea pentru soția lui a rămas la fel de mare, chiar dacă, uneori, nu-și aduce aminte că a întâlnit-o – se căsătoriseră doar cu

Scanned with OKEN Scanner

un an înainte de boala lui. Iar când cântă la pian, cu vocea, sau dirijează, muzica izvorăște din el natural, la fel de fluent ca întotdeauna, făcându-l să pară din nou un întreg.

Pentru Clive, aceste momente de iubire și muzică sunt transformaționale și răsplătitoare. Într-un articol caracteristic evocator din New Yorker, Oliver Sacks descrie situația astfel: „El nu mai are o narațiune interioară; nu-și duce viața în același fel ca noi, ceilalți. E suficient să se așeze în fața clapelor sau să fie lângă Deborah, pentru ca să redevină viu și el însuși".

În ciuda acestor momente de grație, situația lui Clive este, fără îndoială, tragică. Distrugerea sinelui său narativ este mai mult decât un deficit de memorie; este incapacitatea de a se percepe pe sine ca fiind continuu în timp, care vine însoțită și de o erodare a simțului fundamental al identității sale personale – pe care noi îl considerăm în mod natural implicit. Memoria nu este esența sinelui dar, așa cum am aflat din această poveste, sau cum mulți dintre noi știu deja din familie sau de la prieteni care s-au lovit de Alzheimer sau demență, persistența și continuitatea autopercepției sunt aproape imposibile în lipsa ei.

Puterea iubirii dintre Clive și Deborah, care poate restabili simțul identității lui Clive, ne conduce la „sinele social".

Oamenii, la fel ca multe alte animale, sunt ființe sociale. Percepția stării mentale a celuilalt este o capacitate esențială pentru creaturile sociale în orice tip de context și în toate formele de societăți. Se crede adesea despre această capacitate – numită, uneori, „teoria minții" – că se dezvoltă mai degrabă lent la oameni, dar ajunge să joace un rol fundamental pentru majoritatea dintre noi de-a lungul vieții.

Scanned with OKEN Scanner

Uneori, putem fi foarte conștienți de asta, de exemplu atunci când ne întrebăm ce anume gândește despre noi partenerul, prietenul, colegul. Dar chiar și când nu suntem măcinați de interacțiunile noastre sociale, capacitatea noastră de a percepe intențiile, convingerile și dorințele altora rulează mereu în fundal, ghidându-ne comportamentul și modelându-ne emoțiile.

Există o vastă literatură pe tema percepției sociale și a teoriei minții, cuprinzând psihologie, sociologie și, mai recent, domeniul în expansiune al neuroștiinței sociale. Mare parte a acestei literaturi examinează respectivele subiecte în termenii importanței pe care o au în ghidarea interacțiunilor sociale. Aici, aș vrea să întorc lentila invers, pentru a discuta cum anume experiența de *a fi eu* depinde, în mod substanțial, de cum anume percep eu că mă percep alții.

Percepția socială – percepția stărilor mentale ale altora – nu se referă doar la judecata explicită sau „gândirea despre" ce anume pot gândi sau nu alții. Mare parte din percepția noastră socială este automată și directă. Alcătuim percepții despre convingerile, emoțiile și intențiile altora la fel de natural și de ușor cum alcătuim percepții despre pisici sau căni de cafea sau scaune și chiar despre propriul nostru corp. Atunci când îmi torn încă un pahar de vin și observ că prietena mea își aduce paharul gol mai aproape, nu am nevoie să gândesc rațional despre intenția ei; percep pur și simplu că mai vrea și ea vin și că ar fi trebuit să-i torn ei mai întâi. Percep aceste stări mentale la fel de ușor, deși nu neapărat cu aceeași acuratețe, cum percep paharul în sine.

Cum se întâmplă asta? Cred că răspunsul se potrivește, din nou, cu ideea de creier ca mașină de prezis și cu percepția ca proces de inferență a cauzelor semnalelor senzoriale.

Ambele tipuri de percepție – non-socială și socială – implică cea mai bună estimare a creierului referitoare la cauzele input-urilor senzoriale. Uneori, după cum știm

Scanned with OKEN Scanner

cu toții, putem înțelege lucrurile total aiurea atunci când percepem ce e în mintea altcuiva, deși nu confundăm niciodată un pahar de vin cu o mașină (firește, dacă nu halucinăm). Unul dintre motivele ambiguității inerente a percepției sociale este acela că motivele relevante sunt ascunse mult mai adânc. Undele luminoase care declanșează percepția unui pahar de vin provin, mai mult sau mai puțin direct, de la paharul în sine, dar semnalele senzoriale relevante pentru stările mentale ale altora trebuie să treacă prin mai multe etape intermediare – expresii faciale, gesturi, acte de vorbire –, fiecare etapă creând noi oportunități pentru o inferență alăturea cu scopul.

La fel ca în cazul percepției vizuale, cea socială depinde de context și de așteptări și putem încerca să minimizăm „erorile de predicție socială" schimbând datele senzoriale – o formă interpersonală de inferență activă –, precum și prin actualizarea predicțiilor. Inferența activă în percepția socială echivalează cu a ne comporta în așa fel încât să schimbăm starea mentală a celuilalt în așa fel încât să ajungă în conformitate cu ceea ce prezicem că se va întâmpla (sau ne dorim să se întâmple). De exemplu, zâmbim nu numai pentru a ne exprima bucuria, dar și pentru a schimba felul în care se simte partenerul nostru, iar atunci când vorbim, încercăm să inserăm gânduri în mintea celuilalt.

Aceste idei despre percepția socială pot fi legate de sinele social în următorul mod. Capacitatea de a infera stările mentale ale altora necesită, la fel ca orice inferență perceptuală, un model generativ. Așa cum știm, modelele generative sunt capabile să genereze semnalele senzoriale corespunzătoare unei anumite ipoteze perceptuale. Pentru percepția socială, asta înseamnă o ipoteză despre stările mentale ale celuilalt, ceea ce implică un grad înalt de reciprocitate. Modelul meu cel mai bun pentru stările tale mentale va include un model despre maniera în care tu modelezi stările mele mentale. Cu alte cuvinte, pot

Scanned with OKEN Scanner

înțelege ce e în mintea ta doar dacă încerc să înțeleg cum anume percepi tu conținutul minții mele. Doar în acest fel ne putem percepe pe noi înșine, refractați prin mințile celorlalți. Despre asta e vorba atunci când ne referim la sinele social, iar aceste percepții predictive ierarhizate social sunt o parte importantă a întregii experiențe de a fi un sine uman.

O consecință contrariantă a acestui concept de sine social este următoarea: conștiența de sine – nivelul cel mai înalt al sinelui, alcătuit atât din aspecte narative, cât și sociale – ar putea necesita context social. Dacă exiști într-o lume fără alte minți – mai specific, fără nicio altă minte *relevantă* – creierul tău nu va avea nevoie să prezică stările mentale ale altora și, deci, nici să infere că propriile lui experiențe și acțiuni aparțin vreunui sine. Cugetarea de secol XVII a lui John Donne, cum că „nimeni nu e o insulă", ar putea fi literalmente adevărată.

Ești aceeași persoană care erai ieri? Sau, mai bine: te trăiești pe tine în același mod în care o făceai ieri? În lipsa vreunui incident major între timp, vei răspunde, probabil, da. Dar dacă ne gândim la săptămâna trecută, luna trecută, anul trecut, acum zece ani, când aveai patru ani; sau cum va fi când vei avea 94 de ani – vei fi aceeași persoană? Ți se va *părea* așa?

Un aspect remarcabil, dar adesea trecut cu vederea al sinelui conștient este acela că ne percepem pe noi înșine ca fiind continui și unificați de-a lungul timpului. Putem numi asta *stabilitatea subiectivă a sinelui*. Se susține nu numai în termeni de continuitate a memoriei autobiografice, ci și într-un simț mai profund despre noi ca persistând de la o clipă la alta, fie la nivel de corp biologic, fie de identitate personală.

Scanned with OKEN Scanner

Comparativ cu experiențele perceptuale ale lumii exterioare, cele referitoare la sine sunt remarcabil de stabile. Percepțiile noastre despre lume se schimbă permanent, obiecte și scene succedându-se într-un flux continuu de evenimente. Experiențele referitoare la sine par să se schimbe mult mai puțin. Chiar dacă *știm* că ne schimbăm în timp – majoritatea avem, pentru asta, mai multe dovezi fotografice decât e necesar – tot ni se pare că nu ne-am schimbat chiar atât de mult. Doar dacă nu avem vreo boală psihiatrică sau neurologică, experiența de a fi un sine pare să fie un punct stabil într-o lume în schimbare. William James – pionierul de secol XIX din domeniul psihologiei – formulează foarte bine: „spre deosebire de un obiect, care poate fi perceput din diferite perspective sau chiar să nu mai fie perceput deloc, noi avem «aceleași senzații ale aceluiași corp cunoscut, mereu acolo»".

Acum ai putea să consideri că nu e nimic de văzut aici. Până la urmă, corpurile – și alte ținte ale percepției legate de sine – chiar se schimbă mai puțin decât lucrurile pe care le percepem acolo, afară. Mă pot deplasa dintr-o cameră în alta, dar corpul meu și acțiunile mele și perspectiva la persoana I mă vor însoți mereu. Pe aceste baze, nu este surprinzător că sinele este simțit ca schimbându-se mai puțin decât exteriorul. Dar cred că mai e ceva de spus.

Așa cum am văzut în capitolul 6, experiența schimbării este, în sine, o inferență perceptuală. Percepțiile noastre se pot schimba, dar asta nu înseamnă că le și percepem ca schimbându-se. Această diferență este exemplificată de fenomenul „orbirii la schimbare", în care lucruri care se schimbă încet (în afara noastră) nu determină vreo schimbare corespunzătoare la noi. Același principiu se aplică și autopercepției. Devenim alții în fiecare moment. Percepțiile sinelui nostru se schimbă continuu – ești ușor diferit acum de cel care a început să citească acest capitol – dar asta nu înseamnă că și percepem aceste schimbări.

Scanned with OKEN Scanner

Această orbire subiectivă față de sinele în schimbare are consecințe. În primul rând, alimentează ideea falsă că sinele este, mai degrabă, o entitate imuabilă, decât o sumă de percepții. Numai că nu acesta este motivul pentru care evoluția a creionat experiențele sinelui în acest fel. Cred că stabilitatea subiectivă a sinelui trece dincolo de orbirea față de schimbare, garantată de corpurile noastre în schimbare lentă. Trăim cu o formă extremă, exagerată de orbire-față-de-schimbarea-sinelui și, pentru a înțelege de ce, trebuie să înțelegem motivul pentru care ne percepem pe noi mai întâi.

Nu ne percepem pe noi înșine pentru a ne cunoaște pe noi înșine, ne percepem pe noi înșine pentru a ne controla pe noi înșine.

Scanned with OKEN Scanner

9
A fi un animal-mașină

Nu vedem lucrurile așa cum sunt ele,
Le vedem așa cum suntem noi.

ANAÏS NIN

Percepția sinelui nu se referă la descoperirea a ceea ce se află în exteriorul nostru sau înăuntru, în corp. Este despre control fiziologic și reglare – este despre a rămâne în viață. Pentru a înțelege de ce e așa, precum și ce anume înseamnă pentru *toate* experiențele noastre conștiente, să începem prin a analiza o dezbatere foarte veche despre legătura dintre viu și minte.

Pe Scara Vieții – ierarhia medievală creștină despre tot și toate – Dumnezeu e în vârf. Imediat dedesubt sunt ființele angelice, apoi oamenii – împărțiți pe categorii acceptate social – apoi alte animale, plantele și, cel mai jos, mineralele. Totul e la locul său și deține propriile puteri și capacități, determinate de locul pe care îl are pe această scară.

Pe Scară, noi, oamenii, suntem situați, ciudat, la granița dintre lumea spirituală, populată de Dumnezeu și de îngeri, și lumea materială a animalelor, plantelor și animalelor. Avem suflete nemuritoare, fiind capabili de rațiune, iubire și imaginație, dar suntem și pradă pornirilor fizice – durere, foame, instincte sexuale – deoarece suntem legați de un corp fizic.

Timp de secole, mai ales în Europa, Scara Vieții – sau *Scala Naturae* – a oferit un model stabil prin intermediul căruia oamenii și-au putut înțelege locul în natură, precum și valoarea, în comparație cu alți oameni – regele fiind mai

Scanned with OKEN Scanner

sus, comparativ cu țăranul, de exemplu. Apoi, în secolul al XVII-lea, René Descartes a eliminat multele ramificații ale Scării împărțind universul în doar două moduri de manifestare: *res cogitans* (lucruri care țin de minte) și *res extensa* (lucruri care țin de materie).

Această simplificare extinsă a tabloului naturii a venit cu nenumărate alte probleme. Una dintre ele, de ordin metafizic, se referă la cum anume pot interacționa cele două domenii – întrebare care a dus la apariția cercetărilor despre conștiență, cu ale lor bune și cu rele, în mare parte cu rele, încă de pe atunci. Au apărut perturbări în ordinea detaliată de care depinde autoritatea politică și religioasă. În cazul în care animalele aveau elemente de *res cogitans*, orice urmă de rațiune, ce le-ar împiedica să aspire la lumea spirituală, asemeni oamenilor? Iar orice tentativă de a investiga intelectual sufletul – așa cum pleda Descartes – era menită să irite puternica biserică catolică.

Descartes și-a jucat mereu cărțile cu mare grijă în relația cu biserica, mergând până în punctul în care voia să demonstreze existența unui Dumnezeu binevoitor în ale sale *Meditații* – a treia și a cincea. Când vine vorba de animale non-umane, se afirmă adesea că el credea despre ele că erau complet lipsite de conștiență. Deși nu putem fi siguri că e așa, e posibil ca vederile lui să nu fie, de fapt, acestea.* Principala afirmație a lui Descartes despre animale era că acestea sunt lipsite de suflet și de toate atributele care țin de rațional, spiritual și conștiență asociate cu deținerea unui suflet. Istoricul Wallace Shugg sintetizează ideile lui despre materie astfel:

> Corpul – atât al omului, cât și al animalului – ... este doar o mașină de respirat, digerat, perceput și mișcat prin intermediul părților sale, aranjate într-un anume fel.

* Descartes avea un câine numit Monsieur Grat (Domnul Zgârietură) căruia îi era, se pare, foarte devotat. Asta nu-l împiedica, însă, să facă vivisecție pe iepuri.

Scanned with OKEN Scanner

Dar doar la om raţiunea dirijează mişcările corpului, în aşa fel încât să facă faţă neprevăzutului; doar omul îşi dovedeşte raţiunea prin folosirea vorbirii adevărate. Fără minţi care să dirijeze mişcările corpului sau care să recepţioneze senzaţia, animalele trebuie privite ca fiind maşini fără gânduri, fără simţuri, care se mişcă precum mecanismul unui ceas.

Din acest punct de vedere, proprietăţile de carne-şi-sânge ale fiinţelor vii – natura lor de organisme – sunt complet şi explicit irelevante pentru prezenţa minţii, a conştienţei sau a sufletului (orice ar fi acesta). Animalele non-umane pot fi gândite ca *betes-machines* – „animale-maşini". În viziunea carteziană, separarea dintre minte şi viu este la fel de decisivă ca cea dintre *res cogitans* şi *res extensa*.

Întărind ideea că omul este special, Descartes a putut să-şi contracareze eventualii duşmani. Deschisese, însă, o uşă periculoasă. Dacă un animal non-uman este un animal-maşină, iar dacă omul este un fel de animal, şi el – până la urmă, omul e făcut din aceleaşi tipuri de carne, sânge, cartilaje şi oase –, atunci facultăţile minţii şi ale raţiunii ar putea deveni explicabile în termeni mecanicişti, fiziologici?

Filosoful francez Julien Offray de La Mettrie, ale cărui scrieri aparţin mijlocului de secol XVIII, gândea cu siguranţă în acest fel. A extins argumentul animal-maşină al lui Descartes la fiinţele umane, susţinând că şi omul este tot o maşină – *l'homme machine* (omul maşină) –, negând, astfel, statutul special, imaterial al sufletului şi punând la îndoială existenţa unui Dumnezeu. La Mettrie nu era genul care să vrea să protejeze autoritatea religioasă, astfel că viaţa sa a devenit rapid mult mai complicată decât a lui Descartes. În 1748, a fost silit să plece din Olanda, pe care o considera căminul lui adoptiv, ca să intre în slujba regelui prusac Frederick, la Berlin, unde a şi murit, trei ani mai târziu, ghiftuindu-se cu pate.

Scanned with OKEN Scanner

Dacă în viziunea carteziană mintea și viul sunt independente, pentru La Mettrie, ele erau profund conectate, în sensul că mintea putea fi considerată o proprietate a viului. Chiar și în prezent au loc discuții pe tema continuității sau discontinuității viului și minții, în termeni de mecanisme și principii care le susțin pe ambele.

Simpatizez, în această dezbatere, cu La Mettrie, dar, în loc să vorbesc în termeni generali despre „minte", prefer să mă concentrez pe conștiență. Astfel, ajungem în inima teoriei „animal-mașină" despre conștiență și sine. *Trăirea conștientă a lumii din jurul nostru și a noastră în această lume are loc cu, prin și datorită corpului nostru viu.* Alcătuirea noastră de animal nu este numai compatibilă cu percepțiile noastre conștiente despre sine și lume. Propunerea mea este următoarea: nu putem înțelege natura și originea acestor trăiri conștiente decât în lumina naturii noastre de creaturi vii.

Sub exprimările stratificate ale sinelui, implicând amintiri din trecut și planuri de viitor, înaintea simțului explicit al identității personale, sub „eu" și chiar înainte de apariția unei perspective la persoana I și a simțului proprietății asupra corpului, există straturi și mai adânci ale sinelui, încă negăsite. Aceste straturi fundație sunt legate intim mai degrabă cu interiorul corpului, decât cu același corp ca obiect al lumii, și cuprind, de la emoții și stări de spirit – ceea ce psihologii numesc experiențe „afective" – la simțul fundamental, inform și mereu prezent al „existenței" pur și simplu ca organism viu, întrupat.

Vom începe explorarea acestor profunzimi cu emoțiile și stările de spirit. Aceste forme de conținut conștient sunt esențiale pentru experiența de a fi un sine întrupat și – la fel ca toate percepțiile – pot fi înțelese și ele ca

Scanned with OKEN Scanner

cele-mai-bune-estimări-bayesiene despre cauzele semnalelor senzoriale. Caracteristica experiențelor afective este originea cauzelor relevante – care se află mai degrabă în corp, decât în afara lui.

Atunci când ne gândim la percepție, avem tendința de a analiza în funcție de manierele diferite în care simțim lumea exterioară – în special prin familiarele văz, auz, gust, atingere, miros. Tipurile de senzații și percepții orientate către exterior se numesc, generic, exterocepție. Percepția interiorului corpului se numește interocepție – „simțul stării fiziologice interioare".*

Semnalele senzoriale interoceptive se transmit, de obicei, de la organele interne – viscerele – către sistemul nervos central și poartă informație despre starea acelor organe, precum și despre funcționarea corpului ca întreg. Semnalele interoceptive se referă la bătăile inimii, presiunea sangvină, diferite aspecte primare ale chimiei sângelui, gradul de tensiune gastrică, la respirație șamd. Aceste semnale circulă printr-o rețea complexă de nervi și regiuni profunde ale trunchiului cerebral și talamusului înainte să ajungă la acele zone ale cortexului specializate în procesarea interoceptivă – în special la cortexul insular.**

Principala proprietate a semnalelor interoceptive este aceea că ele reflectă, într-un fel sau în altul, cât de bine decurge reglarea fiziologică. Cu alte cuvinte, cât de bine își face treaba creierul în ce privește menținerea noastră în viață.

Semnalizarea interoceptivă este conectată de mult cu emoția și cu stările. În 1884, William James și – independent – Carl Lange au susținut că emoțiile nu sunt nici „entități eterne și sacre" ale filosofilor antici, nici

* Între exterocepție și interocepție se află propriocepția, care se referă la percepția poziției și a mișcărilor corpului (vezi capitolul 5). Este important să nu confundăm interocepția și introspecția, care se referă la examinarea interioară a stărilor noastre mentale.

** Cortexul insular se numește astfel deoarece seamănă cu o „insulă" situată în „marea" corticală.

Scanned with OKEN Scanner

cablate/bobinate în circuitele cerebrale prin evoluție, așa cum propusese Darwin cu puțin timp înainte. Emoțiile sunt – în optica celor doi – percepții ale schimbărilor care survin în corp. Nu plângem pentru că suntem triști, suntem triști deoarece percepem că trupul nostru este într-o stare în care are nevoie să plângă. Frica, din acest punct de vedere, este consecința interocepției mai multor răspunsuri ale corpului declanșate de recunoașterea, de către organism, a pericolului din mediu. Pentru James, emoția *este* chiar percepția schimbărilor corporale, pe măsură ce acestea au loc: „ne pare rău deoarece plângem, suntem furioși deoarece lovim, ne e frică deoarece tremurăm și nu invers – plângem, lovim, tremurăm pentru că ne pare rău, suntem furioși sau ne e frică".

Teoria lui James a întâmpinat o mare rezistență, la vremea ei, parțial deoarece subminează convingerea inițială, naivă, de tip cum-par-lucrurile, cum că emoțiile produc răspunsuri corporale și nu invers. *Pare că* frica – pe care o simțim atunci când ne întâlnim cu un grizzly, de exemplu – este cea care face ca inima să bată tare, nivelul adrenalinei să crească, picioarele să o ia la fugă. Însă am aflat, deja, că trebuie să fim sceptici atunci când lucrurile par într-un anume fel, ele fiind, de fapt, altfel – așa că nu ar fi înțelept să respingem optica lui James doar pe baza primei impresii.

O altă preocupare se referă la faptul că stările corpului nu pot fi delimitate suficient una de alta pentru a susține întreaga desfășurare de emoții care ne însoțesc trăirile. În timp ce detaliile acestei preocupări sunt, încă, subiect de dezbatere, în anii 60 a fost formulat un răspuns foarte convingător, sub forma „teoriilor evaluării". Conform acestor teorii, emoțiile sunt mai mult decât simple „recitări" ale schimbărilor stării corpului. Ele depind de o evaluare cognitivă de nivel mai înalt a contextului în care au loc schimbările fiziologice.

Scanned with OKEN Scanner

Teoriile evaluării rezolvă problema domeniului emoțional deoarece fiecare emoție specifică nu mai are acum nevoie de o anume stare corporală dedicată. Două emoții strâns înrudite – de exemplu, apatia și oboseala – pare că se bazează pe aceeași stare corporală. Diferențele dintre emoții provin din interpretări cognitive diferite ale acestei stări corporale împărtășite. Bineînțeles, poate fi la fel de valabil – și bănuiesc că este, probabil, adevărat – faptul că fiecare emoție chiar are o anumită semnătură întrupată, detaliile fine ale caracteristicilor ei fiind, însă, extrem de dificil de detectat.

Cercetarea experimentală favorită a teoriei evaluării provine de la un studiu inventiv, realizat de Donald Dutton și Arthus Aron în 1974, în care o femeie reporter se apropie de bărbații care trec podurile de peste râul Capilano din nordul Vancouverului. Unul dintre poduri, lung de 137 de metri, este suspendat, are balustrade joase și pare poziționat precar deasupra curenților rapizi și puțin adânci. Celălalt, mai scurt și mai solid, e din cedru și se află în amonte, la doar 3 metri deasupra apei. Atunci când femeia discută cu cei care trec podurile, îi roagă să completeze un chestionar și le oferă și numărul ei de telefon, spunându-le că s-ar bucura să răspundă la orice întrebare suplimentară care ar putea să apară ulterior.

Cercetătorii s-au întrebat dacă nu cumva bărbații de pe podul suspendat interpretează eronat starea de tensiune/excitarea fiziologică – determinată de impresia de instabilitate pe care o lasă podul – mai degrabă ca atracție sexuală, decât ca teamă sau anxietate. Dacă e așa, respectivii bărbați o vor suna pe femeie ulterior, poate chiar propunându-i o întâlnire.

Și chiar așa s-a întâmplat. Femeia a primit mai multe apeluri de la bărbații care au trecut podul suspendat decât de la cei care au trecut pe podul mai solid. Dutton și Aron au numit asta „atribuire eronată a emoției/excitării“:

Scanned with OKEN Scanner

tensiunea interioară, indusă de podul care pare nesigur, a fost interpretată eronat de către sistemul cognitiv superior ca fiind chimie sexuală. În sprijinul acestei interpretări a teoriei evaluării (în condiții de heterosexualitate), s-a constatat că, atunci când reporterul era bărbat, tipul de pod nu a avut niciun efect asupra numărului de apeluri ulterioare.*

Acest studiu, făcut acum mai bine de 40 de ani, are slăbiciuni metodologice evidente comparativ cu standardele actuale, mult mai riguroase, însă tot imperfecte. Nu mai menționăm și etica dubioasă. Dar ilustrează foarte bine ideea conform căreia experiențele emoționale depind de modul în care schimbările fiziologice sunt evaluate de procesele cognitive la nivel superior.

Una dintre limitările teoriilor evaluării constă în aceea că presupune o diferențiere foarte clară între ceea ce considerăm sau nu „cognitiv". Se presupune că sistemele perceptive de nivel inferior, „non-cognitiv", „recită" starea fiziologică a corpului, în timp ce sistemele cognitive de nivel înalt „evaluează" respectiva stare prin procese mult mai abstracte, precum interpretarea sensibilă la context. De exemplu, frica apare atunci când este percepută o anumită stare a corpului, evaluată apoi ca „datorându-se apropierii unui urs". Totuși, din nefericire pentru teoriile evaluării, creierul nu separă clar domeniile „cognitiv" de „non-cognitiv".

Prin 2010, pe când grupul meu din Sussex prindea aripi, am început să mă gândesc la această problemă. Învățasem destul despre interocepție de la colegul meu, Hugo Critchley – unul dintre experții de nivel mondial

* În septembrie 2020, am parcurs renumita creastă Sharp Edge din munții Blencathra, în districtul Lake al Angliei. Deși nu e nevoie de echipament de cățărat, traversarea Sharp Edge nu e chiar ușoară. Creasta zimțată e din roci alunecoase, iar de o parte și de alta sunt prăpăstii adânci, astfel că se pot întâmpla și accidente. În acest loc am observat o piatră răsturnată pe care scria, cu cretă, „Maria, te căsătorești cu mine?". Nu m-am putut abține să nu mă întreb dacă autorul știa de experimentul lui Dutton și Aron și nu a profitat de asta.

Scanned with OKEN Scanner

în domeniu – și am realizat că o manieră de a depăși limitările teoriilor evaluării era să aplic principiile percepției predictive și să tratez emoțiile și stările de spirit – și experiențele din zona afectivă, în general – ca fiind tipuri diferite de halucinație controlată.

Am numit această idee *inferență interoceptivă*. Așa cum creierul nu are acces direct la cauzele semnalelor senzoriale exteroceptive, cum ar fi cele din cazul vederii, care vin din afară, tot așa nu are acces nici la cauzele semnalelor senzoriale interoceptive, care vin din interiorul corpului. *Toate* cauzele semnalelor senzoriale, indiferent de care sunt acelea, sunt ascunse mereu și pentru totdeauna în spatele unui văl senzorial. De aceea, interocepția este înțeleasă cel mai bine ca proces de cea-mai-bună-estimare-bayesiană, fix la fel ca percepția exteroceptivă. La fel cum „roșu" este un aspect subiectiv al predicțiilor bazate-pe-creier despre cum anume anumite suprafețe reflectă lumina, emoțiile și stările de spirit sunt aspecte subiective ale predicțiilor despre cauzele semnalelor interoceptive. Sunt forme, dictate intern, de halucinație controlată.*

La fel ca predicțiile vizuale, cele interoceptive operează pe multe scale de timp și spațiu, susținând cele mai bune presupuneri multi-stratificate, sensibile la context și fluide despre cauzele semnalelor interoceptive. Astfel, inferența interoceptivă rezolvă problema gamei emoționale fără a fi nevoie de trasarea unei limite clare între non-cognitiv și cognitiv. Inferența interoceptivă este, deci, mai parcimonioasă decât teoria evaluării deoarece implică doar un proces (cea mai bună estimare/presupunere bayesiană), nu două (percepția non-cognitivă și evaluarea cognitivă) și, din această cauză, se potrivește mult mai bine cu anatomia creierului.

* Deplasându-mă direct de la teoria evaluării la inferența interoceptivă omit foarte multe studii intermediare. Antonio Damasio, în special, a contribuit decisiv, arătând nu numai că emoția și cogniția sunt înrudite, dar și modul în care depind, amândouă, de corp. Independent de Damasio, Lisa Feldman Barrett a formulat propriile idei, foarte asemănătoare, subliniind importanța predicțiilor interoceptive.

Scanned with OKEN Scanner

Inferența interoceptivă e dificil de testat experimental, în parte deoarece semnalele interoceptive sunt mai greu de măsurat și de lucrat comparativ cu cele exteroceptive, precum vederea. O abordare promițătoare explorează posibilitatea de a folosi răspunsurile creierului la bătăile inimii ca semnătură a erorilor de predicție interoceptive.

Frederike Petzschener, specialistul german în neuroștiință, a demonstrat recent că astfel de răspunsuri, numite „potențiale evocate ale bătăilor inimii", sunt modulate de procese precum a fi atent – așa cum a prezis inferența interoceptivă. Este nevoie, însă, de mai multă cercetare pe această temă.

O altă dovadă indirectă vine din experimentele pe tema simțului de proprietate asupra corpului, cum sunt cele pe care le-am descris în capitolul anterior. Într-un studiu din 2013, condus de Keisuke Suzuki, am constat următoarele: corelarea unor instantanee ale „mâinii de cauciuc", din realitatea virtuală, cu bătăile inimii induce un simț al proprietății mai acut asupra acelei mâini, comparativ cu situația în care instantaneele nu sunt corelate cu bătăile inimii. Asta sugerează că simțul de proprietate asupra corpului depinde de integrarea atât a semnalelor exteroceptive, cât și interoceptive. Această metodă, a „sincronizării cardio-vizuale", a fost folosită și de Jane Aspell și colegii ei într-un aranjament experimental de „iluzie a întregului corp", în care subiecților li se arăta silueta virtuală a propriilor corpuri. Și în acest caz au fost raportate identificări mai puternice cu silueta atunci când imaginea acesteia era arătată sincron cu bătăile inimii. În timp ce aceste studii sugerează inferența interoceptivă, este nevoie de mai multă cercetare și aici, în parte deoarece aceste experimente nu au luat în calcul diferențele individuale în sugestibilitate hipnotică – un factor despre care am aflat că este extrem de important în experimentele de testare a simțului proprietății asupra corpului. Ele depind și de

Scanned with OKEN Scanner

cât de conștientă este o persoană de bătăile propriei inimi, caracteristică ce s-a dovedit frustrant de greu de măsurat.

Din perspectiva teoriei animalului-mașină, cea mai importantă implicație a inferenței interoceptive este aceea că experiențele afective nu sunt doar modelate de predicțiile interoceptive ci chiar alcătuite din acestea. Emoțiile și stările de spirit, la fel ca toate percepțiile, vin din interior spre exterior, nu invers. Fie că vorbim de teamă, anxietate, bucurie sau regret – fiecare experiență emoțională are rădăcinile în cea mai bună estimarea perceptivă top-down despre starea corpului (și despre cauzele respectivei stări). Recunoașterea acestui fapt este primul pas esențial către înțelegerea modului în care experiențele de a fi un sine întrupat sunt legate de materialitatea noastră de carne-și-sânge.

Pentru următorul pas, ne punem următoarea întrebare: *la ce folosesc* aceste percepții ale corpului „din interior"? Percepția exteriorului este, evident, utilă pentru ghidarea acțiunii, dar de ce e nevoie ca starea fiziologică internă să facă parte din viața noastră conștientă? Răspunsul la această întrebare ne duce, iarăși, înapoi în istorie, de data aceasta doar până la mijlocul secolului al XX-lea, la amalgamul de informatică (computer science), inteligență artificială, inginerie și biologie numit *cibernetică*.

În 1950, la răsăritul erei computerului, noile discipline ale ciberneticii și inteligenței artificiale (AI) erau la fel de promițătoare și, în multe feluri, inseparabile. Cibernetica – din *kybernetes* care, în greacă, înseamnă „conducător" sau „guvernator" – a fost descrisă de unul dintre fondatorii ei, matematicianul Norbert Wiener, drept „studiul științific al controlului și comunicării la animale și la mașini". Cibernetica era concentrată mai ales pe control, aplicațiile

Scanned with OKEN Scanner

ei primare fiind în sisteme – cum sunt rachetele ghidate – ce implicau bucle închise de feedback de la output la input. O caracteristică evidentă a acestei abordări era aceea că astfel de sisteme păreau să aibă „scopuri" sau „obiective", cum ar fi lovirea unei ținte.

Acest mod de a gândi despre mașini – ca având „scopuri" – a furnizat o nouă punte dinspre non-viu către viu. Anterior, se credea că doar sistemele biologice pot avea obiective, că se pot comporta conform unui scop interior.* Cibernetica a arătat că nu e așa, subliniind legăturile strânse dintre mașini și animale. În parte din această cauză, s-a separat de alte abordări din AI care subliniau judecata abstractă, des-trupată, offline care avea să fie exemplificată de computerele care jucau șah. Aceste abordări alternative au devenit, atunci, eroii zilei, dominând titlurile de presă și agențiile de finanțare, în timp ce cibernetica a rămas pe margine. Chiar și în relativă obscuritate, cibernetica a furnizat numeroase introspecții valoroase – a căror semnificație începe să fie recunoscută abia acum.

Una dintre aceste introspecții se află într-un articol din 1970, semnat de William Ross Ashby și de Roger Conant, care descrie a lor așa-numită „teoremă a bunului reglator". Conceptul este împachetat frumos în titlul articolului: „Fiecare bun reglator al unui sistem trebuie să fie un model al acelui sistem".

Gândește-te la sistemul centralizat de încălzire sau – la fel de bine – la sistemul de aer condiționat. Să zicem că sistemul este proiectat să mențină temperatura în interior la 19°C. Majoritatea sistemelor de încălzire centrală funcționează pe baza unui feedback simplu: dacă temperatura e prea mică, pornesc, altminteri, se opresc. Să numim acest tip simplu de sistem „Sistemul A".

* Vorbesc în special despre perspectivele post-Iluminism. Sisteme de credințe anterioare, precum Animismul, au atribuit scop (viu și spirit) mult mai liber.

Scanned with OKEN Scanner

Să ne imaginăm acum un sistem mai avansat, „Sistemul B". Acesta e capabil să prezică modul în care temperatura din casă va răspunde la poziția pornit sau oprit a căldurii. Predicțiile se bazează pe proprietățile casei – cât de mari sunt camerele, unde se află radiatoarele, materialul din care sunt făcuți pereții – precum și pe condițiile atmosferice din exterior. Sistemul B ajustează, corespunzător, output-ul boilerului.

Mulțumită acestor capacități avansate, Sistemul B este mai bun decât Sistemul A la menținerea unei temperaturi constante în casă, mai ales dacă acea casă – sau vremea – sunt complexe. Sistemul B este mai bun deoarece are un *model* al casei, care permite prezicerea modului în care temperatura din interior va răspunde la acțiunile pe care le poate întreprinde. Un Sistem B top-end poate fi capabil chiar să anticipeze provocările viitoare legate de temperatură – poate o zi rece la orizont – și să modifice output-ul boilerului în avans, pentru a contracara scăderea temperaturii în casă. Așa cum spun Conant și Ashby, *fiecare reglator al sistemului trebuie să fie un model pentru acel sistem.**

Să mergem cu exemplul ceva mai departe. Să ne imaginăm că Sistemul B a fost dotat cu senzori de temperatură „zgomotoși" imperfecți, care reflectă doar indirect temperatura ambientală a casei. Asta înseamnă că temperatura propriu-zisă nu poate fi „recitată" direct de la senzori; în schimb, trebuie să fie inferată pe baza datelor senzoriale și a așteptărilor anterioare. Sistemul B trebuie să aibă acum un model al (i) cum anume citirile senzorilor săi se înrudesc cu cauzele lor ascunse (temperatura reală a casei) și (ii) cum anume vor răspunde aceste cauze la

* Ne-am putea întreba dacă există vreo diferență între „a fi" un model și „a avea" un model. Cred că putem spune despre sistemele care dețin modele generative explicite, capabile să facă predicții condiționale și contraintuitive, așa cum e Sistemul B, că „au" modele. Reglatorii care sunt relativ ficși și inflexibili, așa cum este termostatul cu feedback simplu, al Sistemului A, pot doar „să fie" un model.

Scanned with OKEN Scanner

diferite acțiuni, cum ar fi ajustarea boilerului sau output-ul radiatorului.

Suntem acum în poziția de a conecta aceste idei despre reglare cu ceea ce știm despre percepția predictivă. Sistemul B lucrează prin inferența temperaturii ambientale din citirile senzorilor, la fel cum creierul nostru face cele mai bune estimări despre cauzele semnalelor sale senzoriale pentru a infera stări ale exteriorului (și ale corpului) și despre cum anume se schimbă în timp. Dar scopul Sistemului B nu este de a-și da seama de „ce e acolo" – în acest caz, temperatura ambientală. Scopul este de a *regla* această cauză ascunsă inferată, de a *acționa* în așa fel încât să păstreze temperatura într-un domeniu confortabil și, ideal, la o valoare unică fixă. Percepția, în acest context, nu se referă la elucidarea a ce anume se află acolo, ci la control și reglare.

Percepția orientată-spre-control – implementată de Sistemul B – este, astfel, o formă de *inferență activă*, procesul prin care erorile de predicție senzorială sunt minimizate mai degrabă prin acțiuni, decât prin actualizarea predicțiilor. Așa cum am explicat în capitolul 5, inferența activă depinde atât de modelele generative, care sunt capabile să prezică modul în care semnalele senzoriale răspund la diferite acțiuni, dar și de modularea echilibrului dintre predicțiile top-down și erorile de predicție bottom-up, astfel încât predicțiile perceptuale să devină autosuficiente.

Inferența activă ne arată că percepțiile predictive pot fi orientate fie către inferența caracteristicilor exterioare (sau ale corpului), fie către reglarea acestor caracteristici – poate fi despre găsirea unor lucruri sau despre controlul lor. Cibernetica vine cu ideea că, pentru unele sisteme, controlul e pe primul loc. Din perspectiva teoremei bunului reglator, întregul aparat de percepție predictivă și de inferență activă provine dintr-o cerință fundamentală cu privire la ceea ce este necesar pentru a reglementa un sistem în mod adecvat.

Scanned with OKEN Scanner

Pentru a răspunde la întrebarea la ce folosesc emoțiile și stările de spirit, avem nevoie de încă un concept din cibernetică – cel de *variabilă esențială*. Introduse tot de Ross Ashby, variabilele esențiale sunt cantități fiziologice, cum ar fi, de exemplu, temperatura corpului, nivelul de glucoză sau de oxigen, și care trebuie menținute mai degrabă în niște limite stricte pentru ca organismul să rămână în viață. Prin analogie, temperatura dorită într-o cameră ar fi „variabila esențială" în cazul unui sistem de încălzire centrală.

Punând toate aceste piese împreună, emoțiile și stările de spirit pot fi acum înțelese drept *percepții-orientate-spre-control, care reglează variabilele esențiale ale corpului.* La asta *folosesc.* Frica pe care o simt atunci când un urs se apropie de mine este o percepție orientată-spre-control a corpului meu – mai specific, „a corpului meu în prezența unui urs care se apropie" – care setează cursul acțiunilor cel mai bine prezise pentru a-mi menține variabilele acolo unde trebuie ele să fie. Important, aceste acțiuni pot fi atât mișcări externe ale corpului – cum ar fi alergarea –, cât și „intero-acțiuni" interne, cum ar fi creșterea ratei cardiace sau dilatarea vaselor de sânge.

Această perspectivă asupra emoțiilor și stărilor le leagă pe acestea și mai strâns de natura noastră de carne-și-sânge. Aceste forme de autopercepție nu se referă doar la înregistrarea stării corpului, fie din afară, fie din interior. Ele sunt legate intim și cauzal de starea noastră de bine și de cât de bine vom fi în viitor, în această întreprindere care e supraviețuirea noastră.

Esențial, făcând această diferență, găsim și motivul pentru care emoțiile și stările de spirit au caracteristicile lor fenomenologice. Experiențe precum frica, gelozia, bucuria, mândria sunt foarte diferite, dar mult mai asemănătoare între ele decât oricare dintre ele cu o experiență vizuală sau auditivă. De ce? Natura experienței perceptuale depinde nu

Scanned with OKEN Scanner

numai de ținta predicțiilor corespunzătoare – poate o cană de cafea pe masă sau o inimă trepidând –, dar și de tipul de predicție care se face. Predicțiile orientate spre găsirea de obiecte vor avea o altă fenomenologie comparativ cu cele orientate spre controlul obiectelor.

Atunci când privesc la cana de cafea de pe birou, am o puternică impresie perceptivă de obiect tridimensional, care există independent de mine. Aceasta este fenomenologia „obiectualității", despre care am vorbit în capitolul 6. Acolo, am propus ideea că obiectualitatea apare în experiența vizuală atunci când creierul face predicții condiționale despre cum anume se vor schimba semnalele vizuale, ca urmare a unei acțiuni sau a alteia – cum ar fi răsucirea unei căni pentru a-i vedea spatele. În acest caz, predicțiile perceptive sunt orientate spre găsirea a ceea ce e acolo, acțiunile relevante fiind cele despre care prezicem, la fel ca în cazul răsucirii, că dezvăluie mai mult despre natura ascunsă a semnalelor senzoriale.

Să considerăm acum un exemplu mai activ: prinderea unei mingi de cricket. Ai putea gândi că cel mai bun mod de a face asta ar fi să-ți dai seama unde se îndreaptă mingea spre pământ și să fugi într-acolo cât de repede poți. Numai că „să-ți dai seama ce e acolo" nu este, de fapt, o strategie bună și nu este ceea ce fac experții în acest sport. În schimb, trebuie să rămâi în mișcare astfel încât mingea „să pară mereu la fel" într-un anume mod – mai specific, astfel încât unghiul elevației privirii tale către minge să crească, dar cu o viteză care descrește constant. Se dovedește că dacă urmezi această strategie – psihologii o numesc „anularea optică a accelerației" – cu siguranță vei intercepta mingea.*

Acest exemplu readuce subiectul controlului. Acțiunile tale – și predicțiile creierului tău despre consecințele lor senzoriale – nu sunt despre a afla unde *este* mingea. Sunt orientate către controlul a cum anume *apare* mingea,

* Dacă ții cont literalmente de acest sfat, mingea te va lovi între ochi.

Scanned with OKEN Scanner

perceptiv. În consecință, experiența ta perceptivă nu va dezvălui poziționarea precisă a mingii în aer, ci ceva asemănător cu „posibilitatea de a o prinde" în timp ce alergi către ea. Percepția, în această situație, este o halucinație care *ne controlează*, pe cât este și o halucinație *controlată*.

Această idee are un pedigree istoric substanțial. În 1970, psihologul James Gibson a susținut că percepem adesea lumea în termeni de ceea ce el a numit „affordances/ oportunități". În viziunea lui Gibson, acestea sunt, mai degrabă, *oportunități pentru acțiune* – o ușă care poate fi deschisă, o minge care poate fi prinsă – decât reprezentări independente de acțiune ale „felului în care sunt lucrurile". O altă teorie, tot din 1970, dar mai puțin cunoscută decât a lui Gibson, pune și mai mult accent pe control. Conform „teoriei controlului perceptiv" a lui William Power, nu percepem lucrurile pentru ca să ne comportăm, ulterior, într-un anume fel. În schimb, la fel ca în exemplul cu prinderea mingii de cricket, ne comportăm în așa fel încât sfârșim prin a percepe lucruri într-un anume fel. În timp ce aceste teorii timpurii se aflau, conceptual, la anumite standarde și în linie cu ipoteza „acțiunea mai întâi" despre creier, pe care am prezentat-o în capitolul 5, ele nu au mecanismele predictive concrete furnizate de viziunea pe care am propus-o asupra percepției: halucinația controlată – sau controlul halucinației.

Ei s-au concentrat, mai degrabă, pe percepția lumii exterioare, nu pe cea a interiorului corpului.

Anxietatea nu are o parte din spate, tristețea nu are laturi, fericirea nu e pătrată. Percepția corpului „dinspre interior", pe care se construiesc experiențele afective, nu furnizează date despre forma și localizarea diverselor mele organe interne – splina e aici, rinichii, dincolo. Nu există o fenomenologie a obiectualității, ca atunci când privim la o cană de cafea de pe masă, și nici nu există nimic precum mișcarea într-un cadru spațial, ca atunci când prindem o minge de cricket.

Scanned with OKEN Scanner

Percepțiile orientate-spre-control care susțin emoții și stări sunt, toate, despre prezicerea consecințelor acțiunilor care păstrează variabilele esențiale ale corpului acolo unde trebuie. De aceea, în loc să trăim emoțiile ca pe obiecte, trăim situația în general, bună sau proastă, și cum anume va evolua. Fie că stau lângă patul de spital al mamei mele, fie calculez cum să scap de un urs, forma și calitatea experienței mele emoționale sunt așa cum sunt – dezolare, speranță, panică, calm – ca urmare a predicțiilor condiționale pe care le face creierul meu despre modul în care diferite acțiuni ar putea influența starea mea fiziologică actuală și viitoare.

În cele mai profunde straturi ale sinelui, chiar sub emoții și stări, se află o experiență subterană din punct de vedere cognitiv, incipientă, greu de descris, *de a fi un organism viu.* Aici, experiențele sinelui apar sub forma unui sentiment nestructurat de *a fi* pur și simplu. Aici ajungem în centrul teoriei animal-mașină: experimentarea conștientă a lumii din jurul nostru și a noastră înșine în ea au loc *cu*, *prin* și *datorită* corpurilor noastre vii. În acest punct, toate ideile despre percepție și sine pe care le-am expus se așază la locurile lor. Să o luăm pas cu pas, de la început.

Scopul primar al oricărui organism este să rămână în viață. Această afirmație este adevărată aproape prin definiție – un imperativ formulat de evoluție. Toate organismele vii se străduiesc să-și mențină integritatea fiziologică în fața pericolului și a oportunității. Pentru asta există creierul. Motivul pentru care evoluția a înzestrat organismele cu creier nu este acela de a scrie poezie, de a rezolva cuvinte încrucișate sau de a face neuroștiință. Din punct de vedere evolutiv, *rolul* creierului nu este gândirea rațională, comunicarea lingvistică și nici măcar percepția lumii. Motivul fundamental pentru care orice

Scanned with OKEN Scanner

organism este dotat cu un creier – sau cu orice tip de sistem nervos – este acela de a-l menține în viață, prin păstrarea variabilelor fiziologice esențiale în domenii de valori restrânse, compatibile cu supraviețuirea.

Aceste variabile esențiale, a căror reglare eficientă determină starea de viu și perspectivele viitoare ale unui organism, sunt cauzele semnalelor interoceptive. La fel ca toate proprietățile fizice, aceste cauze rămân ascunse în spatele unui văl senzorial. Ca și în cazul lumii exterioare, creierul nu are acces direct la stările fiziologice ale corpului, astfel că aceste stări trebuie inferate prin cea mai bună estimare bayesiană.

La fel ca în cazul tuturor percepțiilor predictive, această cea mai bună estimare se obține printr-un proces de minimizare a erorii de predicție, realizat de către creier. În contextul interocepției, aceasta se numește inferență interoceptivă. Ca și pentru vedere și auz – de fapt, pentru *toate* tipurile de percepție – percepția interoceptivă este un fel de halucinație controlată.

În timp ce inferența perceptuală despre lume este adesea orientată spre găsirea lucrurilor, inferența interoceptivă este, în primul rând, despre controlul lor – despre reglarea fiziologică. Inferența interoceptivă exemplifică inferența activă, în care minimizarea erorilor de predicție se face mai degrabă acționând pentru a împlini predicțiile top-down, decât prin actualizarea predicțiilor în sine (deși și asta se întâmplă). Aceste acțiuni reglatoare pot fi externe, cum ar fi să caut de mâncare, sau interne, cum sunt reflexele gastrice sau modificările temporare ale presiunii sangvine.

Acest tip de control predictiv poate suporta răspunsuri *anticipative*, prin predicții despre stări corporale viitoare și despre dependența lor de o acțiune sau alta. Acest tip de control anticipativ poate fi absolut necesar pentru supraviețuire. De exemplu, consecințele pot fi grave dacă aciditatea sângelui ar depăși limitele înainte de a declanșa

Scanned with OKEN Scanner

un răspuns adecvat. Din nou, acțiunile relevante pot fi externe, interne sau ambele. A fugi de un urs înainte să fii atacat este un exemplu simplu de reglare anticipativă. Creșterea ocazională a presiunii sangvine necesară unei fugi eficiente sau chiar și a te ridica de la birou după ce ai lucrat o vreme reprezintă un răspuns anticipativ intern.

Termenul din fiziologie care descrie acest proces este cel de *allostazie*. Spre deosebire de *homeostazie*, care înseamnă doar tendința către o stare de echilibru, allostazia se referă la procesul prin care este asigurată stabilitatea pe parcursul schimbării. Ne putem gândi la inferența interoceptivă ca fiind despre reglarea allostatică a stării fiziologice a corpului.

Așa cum predicțiile despre semnalele senzoriale vizuale stau la baza experiențelor vizuale, predicțiile interoceptive – fie despre viitor, fie despre aici și acum – stau la baza emoțiilor și stărilor. Aceste experiențe din domeniul afectiv au o fenomenologie caracteristică, motivul fiind natura orientată-spre-control și referitoare-la-corp a predicțiilor perceptive de care depind. Ele controlează halucinațiile și sunt, în egală măsură, halucinații controlate.

Deși bine înrădăcinate în reglarea fiziologică, emoțiile și stările de spirit sunt trăite în mare parte, cel puțin parțial, ca legate de lucruri și situații dincolo de sine, din afara corpului. Atunci când simt frică, de obicei îmi este frică de *ceva*. Dar nivelurile cele mai profunde ale sinelui – sentimentul incipient de „doar a fi" – par să nu țină cont, sub nicio formă, de acești referenți externi. Pentru mine, aceasta este starea primară adevărată a sinelui conștient: o predicție perceptivă informă, orientată-spre-control, despre starea fiziologică actuală și viitoare a corpului în sine. Aici începe „a fi tu" și aici găsim conexiunile cele mai profunde dintre viață și minte, dintre natura noastră de animal-mașină și sinele nostru conștient.

Pasul final, fundamental din teoria animalului-mașină este de a recunoaște că totul decurge din acest punct inițial.

Scanned with OKEN Scanner

Nu suntem animalele-mașină ale lui Descartes, la care viul nu are nicio relevanță pentru minte. Este exact invers. *Toate* percepțiile și experiențele noastre, fie că sunt ale sinelui, fie ale lumii, *toate* controlează halucinații și sunt halucinații controlate dinspre interior spre afară. Au rădăcini adânci în mașinăria predictivă din carne-și-sânge care a evoluat, se dezvoltă și operează în fiecare secundă, mereu în lumina impulsului biologic fundamental de a rămâne în viață.

Suntem, în întregime, animale-mașină conștiente.

La finalul capitolului anterior, am semnalat că, în timp ce percepțiile lumii exterioare vin și pleacă, experiențele sinelui par să fie stabile și continue pe parcursul diferitelor scale de timp. Putem constata acum că această stabilitate subiectivă apare în mod natural din teoria animalului-mașină.

Pentru a regla eficient starea fiziologică a corpului, semnalele anterioare celor interoceptive trebuie să aibă precizie înaltă, în așa fel încât să tindă să devină autosuficiente. Aspectul cheie al inferenței active face ca cele mai bune estimări interoceptive să fie orientate către aceste anterioare – către domeniile dorite (prezise) de viabilitate fiziologică. De exemplu, estimez că temperatura corpului meu este mai degrabă constantă în timp și, de aceea – conform inferenței active –, chiar așa se întâmplă. Din acest motiv, sinele neschimbător, ca experiență, provine direct din nevoia de a avea anterioare precise – predicții puternice – despre stări corporale stabile, în scopul reglării fiziologice. Cu alte cuvinte: cât timp suntem în viață, creierul nu-și va actualiza niciodată convingerea anterioară că se așteaptă să fie viu.*

* Un alt mod de a ne gândi la asta este următorul: semnalele senzoriale interoceptive vor fi sistematic „neasistate" pentru a permite intero-acțiunilor să regleze variabilele esențiale, în același fel în care semnalele senzoriale proprioceptive sunt atenuate în timpul acțiunilor externe – cum am discutat în capitolul 5.

Scanned with OKEN Scanner

Mai mult, dat fiind că „schimbarea“ este, în sine, un aspect al inferenței perceptive, creierul poate atenua așteptările anterioare legate de percepția schimbărilor din starea corpului, pentru a se asigura în continuare că variabilele fiziologice esențiale rămân acolo unde trebuie să fie. Asta implică o formă de „orbire la autoschimbare“ – noțiune introdusă tot în capitolul anterior. Din această perspectivă, chiar dacă starea noastră fiziologică se schimbă, noi am putea să nu percepem că se schimbă.

Punând împreună toate aceste idei, ne percepem pe noi înșine ca stabili în timp mulțumită unei așteptări anterioare autosuficiente cum că starea noastră fiziologică este restricționată la un anumit domeniu și, în parte, și unei așteptări anterioare autosuficiente cum că această stare nu se schimbă. Cu alte cuvinte, reglarea fiziologică eficientă se poate baza pe o *percepție eronată* sistematică a stării interne a corpului ca fiind mai stabilă decât e, de fapt, și ca schimbându-se mai puțin decât o face cu adevărat.

Foarte curios, această propunere poate fi generalizată la alte niveluri, mai înalte, ale sinelui, dincolo de starea fundamentală a integrității fiziologice continue. Vom fi capabili să ne menținem mai bine identitatea fiziologică și psihologică, la fiecare nivel al sinelui, dacă nu (ne așteptăm să) ne percepem pe noi ca schimbându-ne continuu. În fiecare aspect al existenței sinelui, ne percepem pe noi înșine ca stabili în timp deoarece ne percepem pe noi pentru a ne controla, nu pentru a ne cunoaște pe noi înșine.

Complementar acestei stabilități subiective, cei mai mulți dintre noi ne percepem pe noi înșine, cea mai mare parte a timpului, ca fiind „reali“. Poate părea evident, dar să ne amintim din capitolul 6 că trăirea lucrurilor din afara noastră ca „existând cu adevărat“ nu reprezintă dovada accesului perceptiv direct la o realitate obiectivă, ci o proprietate fenomenologică ce trebuie explicată. Acolo am propus ca, pentru a fi folositoare unui organism perceptiv,

Scanned with OKEN Scanner

cele mai bune estimări ale noastre să fie trăite mai degrabă *ca și cum ar exista cu adevărat acolo, afară*, decât așa cum sunt, de fapt – constructe cerebrale.

Aceeași judecată se aplică și în cazul sinelui. Așa cum pare că scaunul din colț este cu *adevărat* roșu sau că *chiar* a trecut un minut de când am început să scriu această propoziție, mașinăria predictivă a percepției, atunci când este îndreptată către interior, face să pară că *chiar* există o esență stabilă a unui „eu" în centrul a tot.

Și în același mod în care fenomenologia de a fi real poate lipsi din percepțiile noastre despre lume, și sinele își poate pierde realitatea. Realitatea trăită (și stabilitatea subiectivă) a sinelui poate să dispară și să reapară pe parcursul unei boli și poate fi atenuată sever, chiar anulată în situația depersonalizării psihiatrice. Exemplele extreme de lipsă-de-realitate-a-sinelui se întâlnesc în cazuri grave de delir, descrise prima dată în 1880 de către neurologul francez Jules Cotard. Sinele întrupat este atât de desprins în cazul sindromului Cotard, încât persoanele respective sunt convinse că nu există sau că au murit deja. Desigur, *experiența* irealității sinelui nu înseamnă că orice esență-a-sinelui s-a ridicat și a plecat. Înseamnă doar că percepțiile orientate-spre-control, asociate cu cele mai profunde straturi ale reglării corporale, chiar au luat-o razna.

În expunerea teoriei animalului-mașină nu pretind că am demonstrat cât de necesar e viul pentru conștiență; că există ceva special despre carne, sânge și intestine – sau neuroni biologici – care face ca doar creaturile alcătuite din așa ceva să aibă experiențe conștiente. Asta poate fi adevărat sau nu. Nimic din ce am spus, cel puțin până acum, nu reprezintă o argumentație solidă pentru

Scanned with OKEN Scanner

niciuna dintre variante și nici nu are intenția să fie. Ceea ce susțin este că, pentru a înțelege de ce experiențele noastre conștiente sunt cum sunt sau cum anume sunt experiențele sinelui și cum se leagă ele de experimentarea lumii, am face bine să luăm în seamă rădăcinile, adânc înfipte în fiziologia viului, ale tuturor percepțiilor.

Gândindu-ne la fundamentele materiale ale conștienței, ajungem iar la problema grea. Teoria animalului-mașină accelerează disoluția acestui mister aparent. Extinzând optica halucinației controlate spre cele mai profunde straturi ale sinelui, prin dezvăluirea experienței sinelui-ca-existând-cu-adevărat ca aspect suplimentar al inferenței perceptive, ideea intuitivă pe care se sprijină problema grea devine tot mai fragilă. În special ideea intuitivă a problemei-grele-prietenoase, conform căreia sinele conștient este, cumva, separat de restul naturii – un observator interior imaterial, cu existență reală, care privește afară, la lumea exterioară, materială – se dovedește a fi expresia încă a unei confuzii dintre cum anume *par* și cum anume *sunt* lucrurile.

Secole în urmă, atunci când Descartes și La Mettrie își formulau opiniile despre relațiile dintre viață și minte, nu problema grea era în discuție, ci existența – sau non-existența – „sufletului". Și – poate surprinzător – găsim ecouri ale sufletului și în povestea animalului-mașină. Acest suflet nu este o particularitate imaterială sau o distilare spirituală a rațiunii. Optica animalului-mașină asupra sinelui, cu ale sale conexiuni intime cu corpul, cu ritmurile persistente ale viului, ne readuce într-un trecut eliberat de frământările unei minți computaționale, înainte de dihotomia carteziană dintre minte și materie, dintre rațiune și non-rațiune. Ceea ce am putea numi „suflet", în această optică, este expresia perceptivă a continuității profunde dintre minte și viu. Ceea ce *există cu adevărat* este acea experiență pe care o avem atunci când

Scanned with OKEN Scanner

întâlnim cele mai profunde niveluri ale sinelui întrupat – acele sentimentele incipiente de „doar a fi". Pare în regulă să numim asta un ecou al sufletului, deoarece reînvie concepte mai vechi despre această noțiune eternă – cum ar fi Atman, în hinduism –, concepte care au contemplat esența noastră primordială mai degrabă ca pe o respirație, decât ca pe un gând.

Nu suntem computere cognitive, suntem mașini care simt.

Scanned with OKEN Scanner

10
Un pește în apă

În septembrie 2007, călătoream din Brighton spre Barcelona pentru a ține o prelegere, la o școală de vară, despre „creier, cogniție și tehnologie". Deși eram foarte bucuros că ajung într-un oraș atât de frumos, îndatoririle de acasă mă împiedicau să ajung la timp pentru a asista la un masterclass de 3 ore al lui Karl Friston, eminent specialist în neuroștiință, despre al său „principiu al energiei libere" și aplicațiile lui în neuroștiință. (Friston a apărut și în capitolul 5, unde am făcut cunoștință cu al lui concept de inferență activă.) Îmi doream din toată ființa să fiu la seminarul lui Friston deoarece ideile lui păreau să surprindă, într-o manieră matematică profundă, deși complicată, unele dintre primele mele gândurile despre percepția predictivă și despre sine.

Resemnat că voi rata prelegerea lui, m-am gândit că voi afla mai multe despre ce s-a întâmplat atunci când voi ajunge acolo. Când am ajuns, mai târziu, în acea seară, la barul de pe terasă, am fost întâmpinat de o mare de chipuri uluite. Karl sărise în avionul de Londra imediat după prelegere, lăsând nedumerire în urma lui. După trei ore de matematică detaliată și de neuroanatomie, majoritatea a rămas mai confuză decât la început.

Parte a problemei se pare că a fost amploarea propunerii. Primul lucru care te surprinde la principiul energiei libere este complexitatea ideii. Aduce împreună concepte, perspective și metode din biologie, fizică, statistică, neuroștiință, inginerie, învățare automată (machine learning) și domenii conexe. Iar aplicațiile ei nu sunt limitate doar

Scanned with OKEN Scanner

la creier. Pentru Friston, principiul energiei libere explică toate caracteristicile sistemelor vii, de la autoorganizarea bacteriei, la detaliile fine ale creierului și sistemului nervos, la forma generală și planul corporal al animalelor, ajungând și pe tărâmul evoluției. Este aproape de o „teorie a tot" din biologie – așa cum a fost propusă până acum. Nu e de mirare că oamenii – inclusiv eu – au fost nelămuriți.

Sărim peste 10 ani. În 2017, colegii mei, Chris Buckley, Simon McGregor, Chang-Sub Kim, și cu mine am reușit să publicăm propriul studiu despre „principiul energiei libere în neuroștiință" în *Journal of Mathematical Psychology*. Ne-a luat aproape 9 ani să facem asta, mult mai mult decât am anticipat, dar mă bucur că am perseverat.

Cel puțin, *cred* că mă bucur deoarece, chiar și după toate acele străduințe, tot rămâne ceva ciudat de insondabil, în final. Pe internet apar regulat postări despre eforturile de a înțelege ideile lui Friston. Este „God help us, let's try to understand Friston on free energy" („Doamne ajută, hai să încercăm să înțelegem ce zice Friston despre energia liberă") a lui Scott Alexander. Există inclusiv un cont-parodie pe Twitter, @FarlKriston, unde apar tot felul de vorbe de duh de tipul „I am, whatever I think I am. If I wasn't, why would I think I am?" („Sunt, indiferent de ceea ce cred că sunt. Dacă n-aș fi, de ce aș gândi că sunt?").

Dar principiul energiei libere are valoare deoarece aparenta lui insondabilitate este acompaniată de o eleganță și o simplitate care sugerează o unitate profundă între viu și minte și, astfel, completează teoria animal-mașină a conștienței în mai multe feluri importante.

Și, așa cum vom vedea, abordat suficient de serios, principiul energiei libere – FEP (free energy principle) – nu este chiar atât de greu de priceput.

Scanned with OKEN Scanner

Să lăsăm la o parte, pentru moment, misterioasa „energie liberă“ și să începem cu o declarație simplă despre ce anume înseamnă a exista – pentru un organism, pentru orice, de fapt.

Pentru ca ceva să existe, e nevoie de o delimitare – o margine – între acel ceva și orice altceva. Fără margini, nu există ceva-uri, nu există nimic.

Această delimitare trebuie să și persiste în timp, deoarece lucrurile/entitățile care există își mențin identitatea de-a lungul timpului. Dacă adaugi o picătură de cerneală într-un pahar cu apă, se va dispersa rapid, colorând apa și pierzându-și identitatea. Dacă, în schimb, adaugi o picătură de ulei, deși uleiul se va împrăștia pe suprafața apei, va rămâne recognoscibil și separat de apă. Picătura de ulei continuă să existe deoarece nu s-a dispersat uniform în apă. După un timp, însă, și ea își va pierde identitatea, în aceeași manieră în care stâncile devin praf. Lucruri precum picăturile de ulei și stâncile există, fără îndoială, deoarece au o identitate care persistă pentru o perioadă de timp – lungă, în cazul stâncilor. Dar nici picăturile de ulei, nici stâncile nu-și păstrează *activ* delimitările, doar se dispersează, suficient de încet pentru ca noi să nu observăm că ele există în tot acest timp.

Sistemele vii sunt diferite. Spre deosebire de exemplele de mai sus, sistemele vii își mențin activ marginile de-a lungul timpului – prin mișcare și, uneori, doar crescând. Ele contribuie activ la a se menține distincte de mediul în care se află, iar acest lucru este o trăsătură fundamentală a ceea ce le face vii. Punctul de plecare pentru FEP îl reprezintă faptul că sistemele vii, prin simplul fapt că există, trebuie să reziste activ dispersiei stărilor lor interne. Până să ajungi o pată terciuită pe jos, nu mai ești viu.*

Acest mod de a gândi despre viu ne reamintește conceptul de *entropie*. În capitolul 2, am introdus entropia

* Există câteva cazuri limită fascinante pe care nu le considerăm ca aparținând viului, dar care își mențin clar identitatea – cum ar fi tornadele sau vârtejurile, de exemplu.

Scanned with OKEN Scanner

ca măsură a dezordinii, diversității, incertitudinii. Cu cât stările unui sistem sunt mai dezordonate – ca în cazul unei picături de cerneală dispersate la întâmplare în apă –, cu atât entropia lui e mai mare. Pentru tine, pentru mine, chiar și pentru o bacterie, stările interne sunt mai puțin dezordonate atunci când suntem vii, decât atunci când ne descompunem într-un terci. A fi viu se traduce prin a fi într-o stare cu *entropie scăzută*.

Problema e următoarea. Conform celei de-a doua legi a termodinamicii din fizică, entropia oricărui sistem fizic izolat crește în timp. Toate sistemele de acest tip se îndreaptă spre dezordine, spre o dispersie a stărilor lor constituente, de-a lungul timpului. A doua lege ne spune că momentele în care materia e organizată, cum e cazul organismelor vii, sunt intrinsec improbabile și instabile și că – pe termen lung – suntem cu toții sortiți pieirii. Cumva, însă, spre deosebire de stânci și picături de cerneală, sistemele vii fentează a doua lege, persistând într-o stare precară de improbabilitate. Ele există în afara echilibrului cu mediul lor înconjurător – și asta înseamnă „a exista", în primul rând.

Conform FEP, pentru ca un sistem viu să reziste presiunii legii a doua a termodinamicii, *trebuie să ocupe stări în care se așteaptă să fie*. Fiind un bun bayesian, folosesc „se așteaptă" cu sens statistic, nu psihologic. Este o idee foarte simplă, aproape banală. Un pește în apă se află într-o stare în care se așteaptă să fie, statistic vorbind, deoarece majoritatea peștilor chiar se află în apă în majoritatea timpului. Un pește în afara apei este de neașteptat din punct de vedere statistic, cu excepția situației în care peștele se descompune. Temperatura corpului meu, de cca 37°C, este tot o stare așteptată statistic, compatibilă cu supraviețuirea mea, cu a nu mă descompune.

Pentru orice sistem viu, starea de „a fi viu" înseamnă căutarea activă a unui set anume de stări care sunt revizuite

Scanned with OKEN Scanner

repetat în timp, fie că acestea sunt temperaturi ale corpului, ritmuri cardiace („variabilele esențiale" fiziologice pe care le-am întâlnit în capitolul anterior) sau organizarea proteinelor complexe și a fluxurilor de energie într-o celulă bacteriană. Sunt stări statistic așteptate, cu entropie scăzută, care asigură supraviețuirea sistemului – tipul de stări așteptate, dată fiind creatura în discuție.*

Important, sistemele vii nu sunt sisteme închise, izolate, ci se află în interacțiune deschisă și continuă cu mediile lor, recoltând resurse, nutrienți și informație. Doar profitând de această deschidere, sistemele vii devin capabile să se angajeze în activitatea însetată-de-energie de a căuta stări așteptate statistic, minimizând entropia și contracarând a doua lege a termodinamicii.

Din perspectiva unui organism, entropia care contează este entropia stărilor sale *senzoriale* – acele stări care îl aduc în contact cu mediul său. Să ne imaginăm un sistem viu foarte simplu, cum este o bacterie unicelulară. Această bacterie are nevoie de un anume nutrient pentru a supraviețui și poate simți concentrația acelui nutrient în imediata sa apropiere, *Așteptând* să simtă concentrațiile mari de acel nutrient și căutând activ astfel de semnale senzoriale prin intermediul mișcărilor ei, acest organism simplu se va menține într-un set de stări care îl definesc ca fiind viu. Cu alte cuvinte, a simți concentrații mari de nutrient este o stare statistic așteptată pentru bacterie, stare pe care o verifică activ.

Conform FEP, tot ce am povestit se aplică la *toate* organismele – nu numai la bacterii. Toate supraviețuiesc datorită minimizării în timp a entropiei lor senzoriale,

* Cum de pot fi stările statistic așteptate și improbabile? Aceasta se întâmplă atunci când un sistem populează doar un repertoriu restrâns sau subset de stări – un așa numit „set atractor" – dintr-un număr mai mare de stări posibile. Setul atractor este statistic așteptat, deoarece aici se găsește sistemul, de obicei, dar este și improbabil, deoarece există mai multe alte stări în afara setului decât în interiorul lui. Există mult mai multe feluri de a te descompune decât de a fi în viață.

Scanned with OKEN Scanner

prin care se asigură că rămân în stările statistic așteptate, compatibile cu supraviețuirea.

Aici ajungem la esența FEP, care formulează următoarea întrebare: cum anume reușesc, în practică, sistemele vii să-și minimizeze entropia senzorială? În mod normal, pentru a minimiza o cantitate, un sistem trebuie să fie capabil să o măsoare. Problema aici e că entropia senzorială nu poate fi detectată direct și nici măsurată. Un sistem nu poate „ști" dacă propriile sale senzații sunt surprinzătoare, doar pe baza senzației în sine. (Iată o analogie: este cifra 6 surprinzătoare? E imposibil de răspuns, fără un context.) De aceea entropia senzorială este foarte diferită de noțiuni precum nivelurile de lumină, sau concentrațiile nutrienților din jur, care por fi detectate direct de un organism prin intermediul simțurilor și folosite pentru a genera un comportament.

Aici intră în scenă *energia liberă*. Nu vă faceți griji de nume, care provine din teoriile termodinamicii de secol XIX.* Pentru ceea ce avem noi nevoie, ne putem gândi la energia liberă ca fiind o cantitate ce aproximează entropia senzorială. Este, de asemenea, o cantitate ce poate fi măsurată de către un organism – deci, ceva ce organismul poate minimiza, ceea ce e esențial.

Mergând mai departe cu FEP, putem afirma acum că organismele se mențin în stări cu entropie joasă care le asigură existența în continuare prin minimizarea activă a acestei cantități măsurabile numite energie liberă. Dar ce anume este energia liberă, din perspectiva organismului? După câteva artificii matematice, se dovedește că energia liberă este, în esență, același lucru cu *eroarea de predicție* senzorială. Atunci când un organism minimizează eroarea

* În termodinamică, energia liberă este cantitatea de energie disponibilă pentru a face lucru mecanic la temperatură constantă. Este „liberă" în sensul că e „disponibilă". Tipul de energie liberă din FEP se numește „energie liberă variațională" – termen care provine din învățarea automată (machine learning) și teoria informației, dar care e strâns înrudit cu echivalentul său termodinamic.

Scanned with OKEN Scanner

de predicție senzorială, ca în cazul procesării predictive și al inferenței active, minimizează și această cantitate mai profundă teoretic care este energia liberă.

O consecință a acestei conexiuni este că FEP acreditează ideea din capitolul anterior conform căreia sistemele vii au – sau sunt – modele pentru mediile lor. (Mai specific, modele ale cauzelor semnalelor lor senzoriale.) Aceasta deoarece, în procesarea predictivă, așa cum am văzut în capitolul 5, modelele sunt necesare pentru a aproviziona predicțiile care, în schimb, definesc erorile de predicție. Conform FEP, un sistem poate decide dacă senzațiile sale sunt (statistic) surprinzătoare doar în virtutea faptului că are sau este un model. (Dacă cifra 6 provine din aruncarea unui zar, atunci poți decide cât de surprinzătoare este.)

Conexiunile profunde dintre FEP și procesarea predictivă capătă un înțeles atrăgător. Intuitiv, minimizând eroarea de predicție prin inferență activă, sistemele vii vor ajunge natural în stările în care se așteaptă – sau își prezic – să fie. Văzute astfel, ideile percepției predictive și halucinației controlate (sau a halucinației care ne controlează) decurg, practic, din tentativa ambițioasă a lui Friston de a explica totul în biologie.

Punând toate acestea împreună, peisajul care apare este cel al unui sistem viu care-și modelează activ mediul extern și intern, astfel încât setul de stări care-l definesc ca fiind viu să fie reevaluat permanent, iar și iar – de la bătăile inimii, secundă de secundă, la comemorarea anuală a zilei mele de naștere. Parafrazându-l pe Friston, perspectiva, dinspre FEP, este cea a unor organisme care se adună și modelează informația senzorială în așa fel încât să maximizeze dovezile senzoriale ale propriei lor existențe. Sau, cum îmi place mie să spun, „Mă prezic pe mine, deci exist".

Este important să semnalăm că minimizarea energiei libere – eroarea de predicție senzorială – nu înseamnă că un sistem viu poate scăpa refugiindu-se într-o cameră

Scanned with OKEN Scanner

întunecată și tăcută unde poate rămâne, zgâindu-se la pereți. Te-ai putea gândi că asta e o strategie ideală, input-urile senzoriale de la mediul extern fiind, în acest caz, foarte predictibile. Dar e departe de a fi ideală. În timp, input-uri senzoriale care semnalează alte lucruri, precum nivelurile glucozei în sânge etc., for începe să devieze de la valorile lor așteptate: ți se va face foame dacă stai închis în cameră pentru mult timp. Entropia senzorială va începe să crească, iar non-existența apare la orizont. Sistemele complexe cum sunt organismele vii trebuie să permită schimbarea anumitor lucruri pentru ca alte lucruri să rămână la fel. Trebuie să ne mișcăm ca să ieșim din pat, să ne pregătim micul dejun, iar presiunea sangvină să crească, în consecință, pentru a preveni leșinul. Este forma anticipativă de control predictiv – allostazia – pe care am menționat-o în capitolul anterior. Minimizarea erorii de predicție senzorială pe termen lung înseamnă să ieșim din camera întunecată sau, cel puțin, să aprindem lumina.

O altă îngrijorare referitoare la FEP este aceea că nu poate fi falsificabilă – adică nu se poate demonstra că ar fi eronată, cu ajutorul datelor experimentale. Este adevărat, dar nu e nici unic pentru FEP și nici neapărat problematic. Cel mai bun mod de a gândi despre FEP este, mai degrabă, ca despre o bucată de filosofie matematică, decât ca o anume teorie, evaluabilă prin testarea ipotezei. Așa cum afirmă colegul meu, Jakob Hohwy, FEP abordează întrebarea „care sunt condițiile necesare posibilității de a exista?“ în același mod, de prime-principii, în care Immanuel Kant o formulează pe a lui: „care sunt condițiile necesare posibilității de a percepe?“. Rolul FEP poate fi înțeles ca motivând și facilitând interpretarea altor teorii, mai specifice; teorii care pot fi infirmate prin experiment. Teoria procesării predictive, de exemplu, poate fi considerată falsă dacă se dovedește că creierul nu folosește erorile de predicție senzorială în procesul

Scanned with OKEN Scanner

percepției. În final, FEP va fi judecată după utilitatea ea, nu după cât este ea de adevărată sau falsă empiric.*

Să recapitulăm principalele etape ale FEP. Pentru ca un organism să rămână în viață, el trebuie să se comporte în așa fel încât să se mențină în acele stări (cu entropie joasă) în care „se așteaptă" să fie. Un pește care înoată deasupra unui recif de corali, după hrană, caută proactiv stările senzoriale așteptate, compatibile cu supraviețuirea lui în continuare. În general, sistemele vii fac asta prin minimizarea aproximării măsurabile a entropiei acestor stări, care este energia liberă. Minimizarea energiei libere necesită ca organismul să aibă sau să fie, un model pentru mediul său (care include și corpul). Organismele-care-minimizează energia liberă folosesc, apoi, aceste modele pentru a reduce diferența dintre semnalele senzoriale prezise și cele reale, prin actualizarea predicțiilor și prin execuția de acțiuni. Într-adevăr, date fiind ipotezele matematice plauzibile, energia liberă se dovedește a fi exact același lucru cu eroarea de predicție. Luate împreună, toate acestea înseamnă că procesarea predictivă și halucinațiile controlate, inferența activă și percepția orientată-spre-control – și teoria animal-mașină, de asemenea – în integralitatea lor, pot fi înțelese prin prisma FEP ca decurgând dintr-o constrângere fundamentală referitoare la ce anume înseamnă să fii viu, să existi.

În caz că această abordare rapidă a FEP pare puțin dezorientantă, revin cu asigurarea că nu e nevoie să înțelegem sau să acceptăm FEP pentru a urmări povestea

* Un alt exemplu de principiu asemănător FEP este „principiul acțiunii staționare" din fizică, al lui Hamilton, care poate fi folosit pentru a deriva (testabil) ecuațiile mișcării și chiar relativitatea generală.

Scanned with OKEN Scanner

halucinației controlate și a animalului-mașină din capitolele anterioare.*

Teoria conform căreia noi trăim lumea și sinele prin mecanisme de percepție predictivă adânc înrădăcinate într-un „impuls de a rămâne în viață" („drive to stay alive") se susține prin ea însăși. Totuși, FEP merită luată în considerare deoarece îmbunătățește teoria animal-mașină, în cel puțin trei moduri semnificative.

În primul rând, FEP fixează teoria animal-mașină în fundația fizicii, în special într-o fizică relevantă în ce privește ce anume înseamnă să fii viu. Impulsul de a rămâne în viață al animalului-mașină reapare în FEP ca un imperativ și mai fundamental de a rămâne în stările statistic așteptate, pentru a contracara a doua lege a termodinamicii. Atunci când o teorie poate fi generalizată și stabilizată astfel, ea devine mai convingătoare, mai integrativă și mai solidă.

În al doilea rând, FEP face teoria animal-mașină mai convingătoare, prin repovestirea ei de la coadă la cap. În capitolele anterioare, am început cu provocarea de a infera, din interiorul craniului nostru, cum anume este lumea exterioară, derulând apoi firul ideilor către interiorul corpului – abordând mai întâi experiențele sinelui sub formă de cele mai bune estimări perceptive și, în final, identificând cele mai profunde dinte aceste experiențe cu percepții orientate-spre-control ale corpului însuși. Cu FEP, e invers. Începem cu simpla afirmație că „lucrurile există" și ne îndreptăm de aici către corp și către lumea exterioară. Sosirea în același loc din două puncte de pornire foarte diferite întărește impresia că povestea subiacentă este coerentă și clarifică paralelele,

* Conceptele și matematica din spatele FEP nu sunt simple nici măcar pentru cei care sunt experți în domeniu. Primele propoziții dintr-un manual de mecanică statistică ne avertizează: „Ludwig Boltzmann, care și-a petrecut mare parte a vieții studiind mecanică statistică, a murit în 1906, de propria mână. Paul Ehrenfest, care a continuat lucrarea, a murit la fel în 1933. Acum e rândul nostru să studiem mecanică statistică."

Scanned with OKEN Scanner

altfel obscure, dintre concepte (de exemplu, energie liberă și eroare de predicție).

În al treilea rând, FEP vine cu un aparat matematic bogat. Acest aparat oferă nenumărate oportunități de a dezvolta mai departe ideile pe care le-am prezentat în capitolele anterioare. Iată un exemplu. Când analizăm matematica FEP în detaliu, descoperim că ceea ce chiar trebuie să fac pentru a rămâne în viață este să minimizez energia liberă în viitor – nu doar aici și acum. Și se dovedește că minimizarea acestei erori de predicție pe termen lung înseamnă că trebuie să caut noi senzații *acum*, pentru a reduce incertitudinea despre ceea ce se va întâmpla *în următorul moment*, dacă fac asta-și-ailaltă. Devin un agent curios, căutător de senzații – nu cineva bucuros că se autoizolează în camera întunecoasă. Matematica FEP ajută la cuantificarea acestui echilibru delicat dintre explorare și exploatare și, în schimb, are implicații în ceea ce percepem, cât timp ceea ce percepem este construit mereu și pretutindeni din predicții pe care le face creierul. Introspecții ca aceasta ne vor permite să facem experimente mai bune, să construim punți explicative mai solide, care vor putea susține greutatea acestor experimente și, puțin câte puțin, punte cu punte, să ne aducă mai aproape de o explicație satisfăcătoare a felului în care mecanismele dau naștere minților.

În același timp, deși promovată drept „teorie a tot", FEP nu este și o teorie a conștienței. Relația dintre FEP și conștiență este aceeași ca relația dintre FEP și teoriile predictive, bayesiene, ale creierului: sunt teorii *pentru* știința conștienței, în sensul problemei reale, nu *ale* conștienței, în sensul problemei grele. FEP vine cu perspective și unelte noi în provocarea de a explica fenomenologia în termeni de mecanism. În schimb, noțiunile de halucinație controlată și animal-mașină aduc, în matematica austeră a FEP, o relevanță nouă pentru conștiență – și la ce poate fi bună o teorie a tot, dacă nu are ceva de spus despre asta?

Scanned with OKEN Scanner

Mulți ani după prima mea întâlnire neliniștitoare cu FEP, am petrecut câteva zile cu Karl Friston – și cu încă vreo 20 de alți specialiști în neuroștiință, filosofi și fizicieni – la o mică adunare pe insula greacă Aegina, situată la o oră de mers cu un feribot din Atena. Era un alt septembrie, de data asta în 2018, la scurt timp după faza delirului mamei mele. La fel ca în cazul excursiei la Barcelona, mai bine de 10 ani înainte, speram să prind și niște soare de vară târzie. Planul era să discutăm despre FEP, cu accent pe relația sa cu teoria informației integrate a conștienței (IIT = integrated information theory) – teoria la fel de ambițioasă pe care am explorat-o în capitolul 3. În loc de soare blând și cer albastru, am fost întâmpinați de o furtună puternică, o așa-numită „medicane" (echivalentul hurricane/taifun, dar în Mediterana, n. trad.), care a aruncat mesele și scaunele în mare și a transformat altminteri placida Mediterană într-o furie înspumată.

În timp ce stăteam în anexa dedicată conferinței, cu ușile trântindu-se în rafale și cu ramuri lovind în ferestre, am realizat brusc cât de extraordinar era să te ocupi de conștiență într-o vreme în care aveam două teorii foarte ambițioase și detaliate matematic, care nu păreau să comunice una cu alta în vreun fel. Lipsa asta de interacțiune putea să pară descurajantă, dar mi s-a părut fascinant să mă aflu în această situație.

Furtuna a continuat toată ziua. Au fost lansate câteva idei, dar impresia mea era că doar încercăm să ne găsim unii pe alții în semiîntuneric. FEP și IIT sunt, ambele, teorii extraordinare, dar în sensuri diferite. FEP pornește de la simpla afirmație că „lucrurile există" și derivă, de aici, întregul conținut al neuroștiinței și biologiei, dar nu și al conștienței. IIT pornește de la simpla afirmație „conștiența există" și, de aici, lansează un asalt direct asupra problemei grele. Nu e surprinzător că vorbesc, adesea, una pe lângă alta.

Scanned with OKEN Scanner

Doi ani mai târziu, pe când finisam această carte, cele două teorii încă trăiau în lumi diferite. Dar acum există câteva tentative care compară predicțiile lor experimentale. Discuțiile pentru planificarea acestor experimente – din care am norocul să fac parte – au fost, pe rând, revelatoare, dar și frustrante, în mare parte ca urmare a diferenței dramatice între punctele de pornire și între scopurile explicative ale fiecăreia dintre teorii. Ce turnură vor lua aceste experimente, rămâne de văzut. Impresia mea este că vom învăța multe lucruri folositoare, dar nici FEP, nici IIT nu va fi considerată o teorie a, sau pentru conștiență.

Propriile mele idei despre halucinațiile controlate și animale-mașină se află undeva între cele două. Împărtășesc, cu FEP, o profundă fundamentare teoretică în natura sinelui și valorifică puternica mașinărie matematică și conceptuală a creierului predictiv. Împărtășesc, cu IIT, accentul pe proprietățile fenomenologice, subiective ale conștienței – deși cu adevărata problemă, nu cu cea grea, în vizor. Decât o confruntare a FEP cu IIT, mai bine le aducem împreună, cu ajutorul teoriei animal-mașină a conștienței și a sinelui, și împletim perspective din ambele teorii, într-o imagine mulțumitoare a de ce suntem ceea ce suntem.

Înapoi pe Aegina, întâlnirea s-a terminat, ca majoritatea, fără sunet de fanfară. La vremea când am luat feribotul către Atena, furtuna se potolise și marea era liniștită. Să vin în această excursie fusese o decizie foarte grea pentru mine. Am pierdut câteva evenimente personale importante în Brighton. În final, stând pe puntea însorită, m-am împăcat cu acea decizie. Am început să mă gândesc cum anume am decis, de ce este mereu atât de greu și, la scurt timp, cum anume ia oricine altcineva decizii și ce anume înseamnă pentru noi să ne simțim în control referitor la alegerile noastre, la comportament în general.

Odată ce începi să te gândești la liber arbitru, nu te mai poți opri.

Scanned with OKEN Scanner

11
Grade de libertate

Își îndoia degetul, apoi îl îndrepta. Misterul era în secunda de dinaintea mișcării, secunda separatoare dintre nemișcare și mișcare, când intenția ei se materializa. Era ca un val care se sparge. Dacă doar ar ajunge pe creasta lui, credea ea, ar fi putut găsi secretul despre ea, despre acea parte din ea care era cu adevărat la comandă. Își aduse degetul arătător mai aproape de față, fixându-l cu privirea, îndemnându-l să se miște. A rămas nemișcat pentru că ea se prefăcea... Și când l-a îndoit, în final, părea că acțiunea a pornit din deget, nu dintr-o parte a minții ei.

IAN McEWAN, *Atonement/Ispășire*

Care este acel aspect din *a fi tu* de care te agăți cel mai puternic? Pentru mulți, este sentimentul de a fi în control în ce privește propriile lor acțiuni, de a fi autorul gândurilor lor. Este ideea convingătoare, dar complexă, după care noi acționăm conform propriei noastre voințe.

Ian McEwan găsește această complexitate chiar și în simpla flexare a unui deget. Tânăra de 13 ani, Briony Tallis, simte că intențiile ei conștiente, de exemplu cea de a îndoi un deget, provoacă acțiuni fizice – îndoirea efectivă a degetului. Linia cauzalității aparente merge direct de la intenția conștientă la acțiunea fizică. Și simte că în acest proces se află esența primordială a sinelui, a ceea ce este ea. Dar când Briony sondează mai adânc în aceste experiențe, lucrurile nu sunt chiar așa de simplu. Unde a început mișcarea? În minte, sau în deget? Cine anume, intenția – sau „sinele" ei – este cea/cel care produce acțiunea sau

Scanned with OKEN Scanner

experiența intenției este rezultatul percepției degetului care începe să se miște?

Briony Tallis intră și ea, astfel, în rândul celor măcinați de astfel de întrebări. Puține subiecte din filosofie și neuroștiință au fost atât de inflamatoare precum liberul arbitru. Ce anume este, dacă există, cum se întâmplă, chiar contează – consensul pe aceste teme a rămas un țel de neatins. Nu e nimic clar nici despre *trăirea* liberului arbitru – este o experiență singulară sau o categorie de experiențe înrudite, diferă de la om la om etc. În mijlocul întregii confuzii persistă, însă, o idee stabilă. Atunci când ne exercităm liberul arbitru, simțim – în cuvintele filosofului Galen Strawson – un sentiment „radical, absolut, de *depinde de mine* în ce privește alegerea și acțiunea". Un sentiment că sinele joacă un rol cauzal în acțiune, într-un fel în care nu se pune problema atunci când ne referim la un răspuns pur reflex, cum este cel de a-ți retrage mâna atunci când te urzici. Din acest motiv, experiențele liberului arbitru însoțesc în mod natural acțiunile voluntare – fie că vorbim de îndoirea unui deget, decizia de a face o cană de ceai sau o schimbare în carieră.

Atunci când fac o acțiune „liber arbitrată", simt, oarecum, că *sinele meu* este cauza acelei acțiuni. Poate mai mult decât orice alt tip de experiență, cea a voinței ne face să simțim că există un *sine* conștient imaterial, care mânuiește cumva lumea materială. Asta e ceea ce pare.

Numai că experiențele voinței nu dezvăluie existența unui sine imaterial, care are o putere cauzală asupra evenimentelor fizice. În schimb, cred că este vorba despre forme diferite de percepție relative la sine. Mai precis, că sunt percepții relative la sine asociate cu acțiuni voluntare. La fel ca toate percepțiile – fie relative la sine, fie la ceea ce e în afara sinelui –, experiențele voinței sunt construite conform principiilor celor mai bune estimări bayesiene și joacă un rol important – chiar esențial – în ghidarea a ceea ce facem.

Scanned with OKEN Scanner

Mai întâi, să stabilim clar ce anume *nu* este liberul arbitru. Nu este o intervenție în fluxul de evenimente fizice din univers, mai precis, din creier, care să aibă drept urmare desfășurarea unor lucruri care, altminteri, nu s-ar fi desfășurat. Acest liber arbitru „ciudat" invocă dualismul cartezian, cere libertate din legile cauzei și efectului și nu oferă nimic cu valoare explicativă în schimb.

Eliminând ciudatul liber arbitru dintre variante, înseamnă că putem renunța și la permanenta, dar eronat direcționata grijă că *determinismul* este sau nu adevărat. În fizică și în filosofie, determinismul este ideea conform căreia toate evenimentele din univers sunt complet determinate de cauze fizice deja existente. Alternativa la determinism este aceea ca șansa să apară în univers de la zero, fie prin fluctuații într-o supă cuantică, fie prin alte principii încă necunoscute ale fizicii. Dacă determinismul contează pentru liberul arbitru – asta este o discuție fără sfârșit. Fostul meu șef, Gerald Edelman, a sintetizat totul printr-o frază provocatoare: indiferent de ceea ce gândești despre liberul arbitru, suntem determinați să-l avem.

Odată ce ciudatul liber arbitru a ieșit din peisaj, putem constata cu ușurință că discuțiile pe marginea determinismului nu contează deloc. Nu mai este nevoie să permitem nicio intervenție non-deterministă. Din perspectiva liberului arbitru ca experiență perceptivă, pur și simplu nu este nevoie de nicio întrerupere în fluxul cauzal al evenimentelor fizice. Un univers determinist se descurcă foarte bine. Iar dacă determinismul este fals, nu se schimbă nimic, pentru că exercitarea liberului arbitru nu înseamnă comportament haotic. Acțiunile voluntare nici nu *par* și nici nu *sunt* haotice.

La începutul anilor 80, la Universitatea California din San Francisco, Benjamin Libet, specialist în neuroștiință,

Scanned with OKEN Scanner

a făcut o serie de experimente – considerate și acum controversate – pe tema bazei cerebrale a acțiunii voluntare. Libet a profitat de un fenomen bine-cunoscut, numit „potențial pre-motor" („readiness potential") – un mic semnal EEG, provenind de undeva din cortexul motor, și care precede mereu acțiunile voluntare. Libet a vrut să știe dacă acest semnal cerebral poate fi identificat nu numai înaintea unei acțiuni voluntare, ci și înainte ca persoana să fie *conștientă de intenția* de a face acțiunea.

Schema lui experimentală, descrisă în imaginea de mai jos, era foarte directă. Libet i-a rugat pe participanți să-și flexeze încheietura dominantă într-un moment ales de ei – pentru a declanșa o acțiune voluntară spontană, la fel ca Briony în romanul lui McEwans. De fiecare dată, măsura timpul precis al mișcării, în timp ce cu EEG înregistra activitatea cerebrală înainte și după declanșarea mișcării. Esențial, și-a rugat voluntarii să estimeze când anume simt „nevoia" să facă fiecare mișcare: momentul exact al intenției conștiente, creasta valului care se sparge. Au făcut asta notându-și poziția unghiulară a unui punct în rotație pe ecranul unui osciloscop în momentul la care simțeau intenția de a face mișcarea, raportând, ulterior, respectiva poziție.

Datele erau clare. Făcând media după numeroase încercări, potențialul pre-motor a fost identificabil cu sute de milisecunde *înainte* de intenția conștientă de a face mișcarea. Cu alte cuvinte, în momentul în care o persoană este conștientă de intenția ei, potențialul pre-motor este deja declanșat.

O interpretare obișnuită a experimentului lui Libet este aceea că „dovedește lipsa liberului arbitru". Într-adevăr, nu e o veste bună pentru ciudatul liber arbitru (nici nu ar fi nevoie de mai multe vești proaste), deoarece pare să excludă posibilitatea ca voința să fi provocat acțiunea voluntară. Libet însuși era suficient de îngrijorat de consecințe, așa

Scanned with OKEN Scanner

că, într-o încercare de salvare care pare disperată acum, a venit cu ideea că între momentul în care s-a simțit nevoia și acțiunea în sine a rămas suficient timp pentru ciudatul liber arbitru să intervină și să *prevină* acțiunea. Dacă nu există o libertate autentică (adică ciudată) de a alege, gândea Libet, măcar există o libertate „de a nu alege". E o șmecherie drăguță, numai că nu ține. Inhibiția conștientă nu e mai miraculoasă decât intenția conștientă originală.

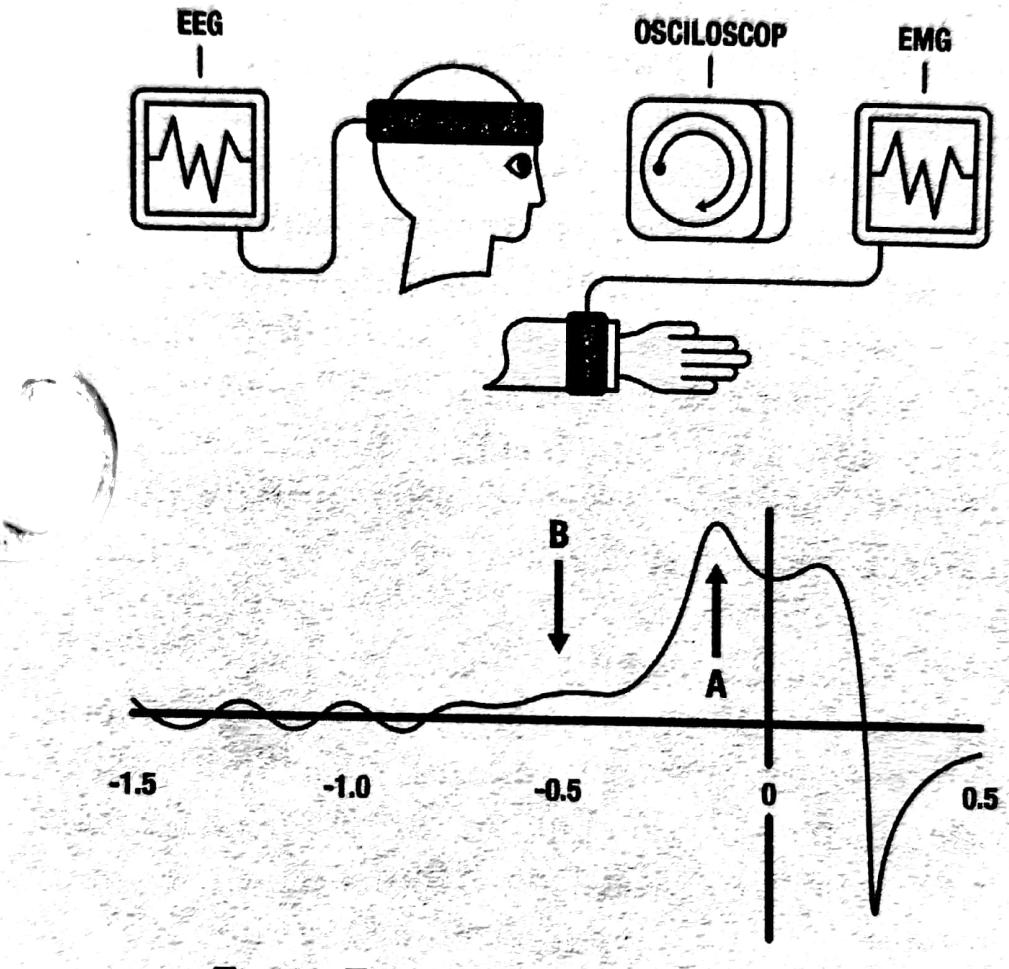

Fig. 19: Experimentul lui Benjamin Libet.*

Ce anume au avut de spus observațiile lui Libet despre liberul arbitru reprezintă subiect de dezbatere de zeci de ani. Este, într-adevăr, straniu că potențialul pre-motor

* Un voluntar este rugat să-și flexeze încheietura într-un moment ales de el, în timp ce notează poziția unui punct pe osciloscop atunci când simte intenția conștientă de a flexa. Alte dispozitive măsoară activitatea mușchilor (EMG) și a creierului (EEG). Graficul reprezintă curba tipică de EEG la începutul mișcării (0 sec). Săgețile arată momentul nevoii conștiente (A) și declanșarea potențialului pre-motor (B).

Scanned with OKEN Scanner

poate fi identificat cu atâta timp înaintea acțiunii voluntare. În termenii creierului, jumătate de secundă este extrem de mult. Abia în 2012 o nouă idee și un experiment inteligent au restabilit lucrurile, atunci când specialistul în neuroștiință Aaron Schurger a realizat că potențialele pre-motorii pot să nu fie semnele inițierii unei acțiuni de către creier, ci artefacte generate de modul în care se face măsurătoarea.

Potențialele pre-motorii sunt măsurate, în mod obișnuit, prin analiza acelor intervale din EEG premergătoare momentului în care s-a produs o acțiune voluntară. Schurger a realizat că, procedând astfel, cercetătorii ignorau toate celelalte intervale în care *nu* se întâmplă acțiuni voluntare. Cum ar arăta EEG în aceste intervale? Poate că activități similare potențialelor pre-motorii au loc permanent, dar nu le vedem pentru că nu ne uităm la ele?

Acest raționament poate fi lămurit printr-o analogie cu mașinăria de măsurat forța de prin parcurile de distracții și pe care, dacă lovești suficient de puternic într-un anume punct cu un ciocan, trimiți în sus un puc de hockey, care face să sune un clopot. Dacă nu lovești suficient de puternic, pucul cade înapoi, fără să atingă clopotul. Dacă am examina științific traiectoria pucului doar după situațiile în care se aude clopotul, am putea concluziona, eronat, că un puc care pleacă în sus (potențialul pre-motor) ajunge întotdeauna la clopot, făcându-l să sune (acțiunea voluntară). Pentru a înțelege cum anume funcționează mașina de măsurat forța, trebuie să examinăm traiectoria pucului și în situațiile în care clopotul *nu* a sunat.

Schurger a atacat această problemă făcând o modificare inteligentă a designului lui Libet – subiecții continuau să execute acțiuni voluntare spontane, dar erau solicitați ocazional, prin intermediul unui bip zgomotos, să execute aceleași acțiuni într-o manieră indusă de stimul, non-voluntară. Descoperirea lui a fost că, atunci când

Scanned with OKEN Scanner

subiecții răspundeau rapid la bip, profilul lor EEG afișa ceva ce arăta ca un potențial pre-motor, extins mult înainte de bip, chiar dacă ei nu pregătiseră vreo acțiune voluntară la momentele respective. Dimpotrivă, profilurile EEG ale celor care răspundeau lent la bip nu prezentau semne de ceva care să semene cu un potențial pre-motor.

Schurger și-a interpretat datele propunând ca potențialul pre-motor să *nu* fie semnalul că creierul inițiază o acțiune voluntară, ci un tipar fluctuant al activității cerebrale care, ocazional, depășește un prag, declanșând o acțiune voluntară. Acesta este motivul pentru care, în experimentul standard al lui Libet, se poate observa o ușoară pantă în profilul EEG atunci când îl analizăm retroactiv de la momentul la care a avut loc acțiunea voluntară. De aceea, atunci când un bip declanșează o acțiune, răspunsul comportamental va fi mai rapid, dacă această fluctuație se apropie de prag, și mai lent, dacă e departe de prag. Aceasta înseamnă, în schimb, că vom vedea ceva ce seamănă cu un potențial pre-motor dacă privim retroactiv de la momentele de răspuns rapid – atunci când activitatea e aproape de prag –, dar nu și de răspuns lent – atunci când activitatea e departe de prag.

Experimentul elegant al lui Schurger explică de ce vedem potențiale pre-motorii atunci când căutăm semnături neuronale ale acțiunilor voluntare și de ce sfârșim în eroare dacă le considerăm ca fiind cauzele specifice ale acestor acțiuni. Atunci cum am putea interpreta aceste tipare fluctuante ale activității cerebrale? Interpretarea mea preferată revine la ideea cu care am început: actele de voință sunt forme de percepție relative la sine. Prin prisma experimentului lui Schurger, potențialul pre-motor seamănă mult cu activitatea asociată cu acumularea, de către creier, a datelor senzoriale în scopul generării celei mai bune estimări bayesiene. Cu alte cuvinte, sunt amprentele neurale ale unui anume tip de halucinație controlată.

Scanned with OKEN Scanner

Tocmai am făcut un ceai.

Să folosim acest exemplu pentru dezvolta ideea conform căreia acțiunile voliționale – și cele voluntare – sunt percepții legate de sine. Trei sunt caracteristicile care definesc majoritatea, dacă nu toate, acțiunile voliționale.

Prima caracteristică este sentimentul că *fac ceea ce vreau să fac*. Fiind englez – măcar pe jumătate –, a face ceai este în acord cu convingerile mele psihologice, cu valorile și aspirațiile, precum și cu starea mea fiziologică de la acest moment și cu oportunitățile – ofertele/affordances – mediului meu. Îmi era sete, aveam și ceai, nimeni nu mă oprea sau mă forța să beau ciocolată caldă, deci am făcut un ceai și l-am băut. (Desigur, dacă sunt forțat să fac ceva „împotriva voinței mele", aș putea să simt că acțiunile mele sunt voluntare la un anumit nivel, dar involuntare la altul.)

Deși prepararea ceaiului a fost în deplin acord cu convingerile mele, cu valorile și cu aspirațiile, nu am ales respectivele convingeri, valori și aspirații. Am vrut o cană de ceai, dar nu am ales să vreau o cană de ceai. Acțiunile voluntare sunt voluntare nu pentru că pogoară de la un suflet imaterial sau pentru că urcă din vreo supă cuantică. Sunt voluntare deoarece exprimă ceea ce eu, ca persoană, vreau să fac, chiar dacă nu pot alege ceea ce vreau. Așa cum spunea filosoful de secol XIX, Arthur Schopenhauer, „omul poate face ceea ce dorește, dar nu-și poate dori ceea ce dorește."

A doua caracteristică definitorie este sentimentul că *aș fi putut face altminteri*. Atunci când simt că acțiunea mea e voluntară, caracterul acestei experiențe nu este doar că eu am făcut X, ci că eu am făcut X și nu Y, chiar dacă aș fi putut face Y.

Am făcut ceai. Puteam face altfel? Într-un fel, da. Este și cafea în bucătărie, deci puteam face cafea. Iar când am făcut ceai, cu siguranță că *mi s-a părut* că puteam să fac cafea, în schimb. Dar nu voiam cafea, voiam ceai și, cât timp

Scanned with OKEN Scanner

nu-mi pot alege dorințele, am făcut ceai. Dată fiind starea precisă a universului la acest moment, care include și starea corpului și a creierului meu, toate având cauze precedente, deterministe sau nu, întinzându-se până la originea mea ca semi-englez băutor de ceai și chiar dincolo, *nu puteam face altfel*. Nu poți relua aceeași înregistrare și să aștepți alte rezultate, în afară de câteva diferențe neinteresante, consecința aleatoriului. Fenomenologia relevantă – *sentimentul* că puteam face altminteri – nu este o fereastră transparentă asupra modului în care cauzalitatea operează în lumea fizică.

A treia caracteristică definitorie este aceea că acțiunile voluntare *par să vină, mai degrabă, din interior* decât să fie impuse de altundeva. Aceasta este diferența dintre acțiunea reflexă, cum ar fi retragerea piciorului când ne lovim la degetul mare, și echivalentul ei voluntar – retragerea piciorului pentru a lovi o minge. Este sentimentul pe care îl avea Briony Tallis atunci când încerca să se prindă pe sine pe creasta valului, gata să se spargă, al intenției ei conștiente de a-și îndoi degetul.

Una peste alta, percepem o acțiune ca fiind voluntară – „la libera alegere" – atunci când inferăm despre cauzele ei că vin mai ales din interior, într-o manieră care rezonează cu convingerile și țelurile fiecăruia dintre noi, rupte de cauzele alternative potențiale din corp sau din afara lui, iar asta sugerează posibilitatea de a fi făcut altminteri. Așa se simt actele voliționale, din interior și, la fel, așa arată din afară acțiunile voluntare.*

Următorul pas este să ne întrebăm cum anume permite și implementează creierul astfel de acțiuni. Aici intră în scenă „gradele de libertate" – titlul capitolului. În inginerie și matematică, un sistem are tot atâtea grade de libertate

* Uneori, acțiunile voluntare sunt trăite ca solicitând efort conștient, sau „voință". Scrierea acestei note de subsol, de exemplu, pare solicitantă. Dar multe acțiuni voluntare autoinițiate cer puțin efort conștient, chiar deloc. De aceea, este important să nu confundăm voința cu liberul arbitru (cu trăirea lui).

Scanned with OKEN Scanner

câte tipuri de a răspunde la diverse stări de lucruri. O piatră nu are, practic, grade de libertate, în timp ce un tren pe o cale ferată unică are un singur grad de libertate (înainte sau înapoi). O furnică poate avea multe grade de libertate dictate de felul în care sistemele ei biologice de control răspund mediului ei, în timp ce tu sau eu avem mult mai multe grade de libertate, ca urmare a complexității extraordinare a corpurilor și creierelor noastre.

Comportamentul voluntar depinde de abilitatea de a controla toate aceste grade de libertate în maniere care rezonează cu convingerile, valorile și scopurile noastre, și care sunt separate adaptativ de exigențele imediate ale mediului și ale corpului. Această abilitate de a controla este implementată de către creier nu printr-o anumită zonă, unde ar rezida „voința", ci printr-o rețea de procese distribuite în multe regiuni ale creierului. Execuția celei mai simple acțiuni voluntare – apăsarea unui buton pentru a porni fierbătorul, flectarea degetului de către Briony – este susținută de o astfel de rețea. Conform specialistului în neuroștiință Patrick Haggard, ne putem gândi la această rețea ca implementând trei procese: un „ce" timpuriu, care specifică acțiunea care se execută, un „când" ulterior, ce determină intervalul acțiunii și un „dacă" final, care permite anularea sau inhibarea acțiunii în ultimul moment.

Componenta „ce" a voinței integrează seturi de convingeri, scopuri și valori, organizate ierarhic, cu percepții ale mediului, pentru a specifica o singură acțiune din nenumăratele posibilități. Întind mâna spre ceainic pentru că mi-e sete, îmi place ceaiul, este momentul potrivit al zilei, ceainicul este la îndemână, nu am vin – șamd. Aceste percepții, convingeri și scopuri ierarhizate implică multe zone diferite ale creierului, mai ales în partea frontală a cortexului. Componenta „când" specifică intervalul de timp al unei anumite acțiuni și este asociată cel mai strâns cu nevoia subiectivă de a face mișcarea – aceeași nevoie

Scanned with OKEN Scanner

care o preocupa pe Briony Tallis și pe care a măsurat-o Benjamin Libet. Baza cerebrală a acestui proces se localizează în aceleași regiuni asociate cu potențialul pre-motor. Într-adevăr, stimularea electrică blândă a acestor regiuni – mai ales a ariei motorii suplimentare – poate genera o nevoie subiectivă de a face mișcarea, chiar și în lipsa oricărei mișcări. Iar componenta finală, „dacă", permite o verificare de ultim moment asupra oportunității de a continua. Atunci când anulăm o acțiune în ultimul moment – poate nu am lapte – intră în acțiune procesul de „inhibiție intențională". Aceste procese inhibitorii sunt și ele localizabile în zone mai frontale ale creierului.

Aceste procese întrețesute se desfășoară într-o buclă continuă, fără început și fără sfârșit, ce include creierul, corpul și mediul, implementând o formă continuă și extrem de flexibilă de comportament orientat spre obiectiv. Această rețea de procese canalizează o gamă largă de cauze potențiale într-un flux unic de acțiuni voluntare – uneori, de inhibare a lor. Iar cea care susține experiențele subiective ale voinței este percepția operării acestei rețele, a buclei care trece prin corp, spre exterior și înapoi.

Mai mult, cât timp acțiunea în sine este o formă de inferență perceptuală autosuficientă, așa cum am văzut în capitolul 5, experiențele perceptive ale voinței și abilitatea de a controla multe grade de libertate sunt cele două fețe ale aceleiași monede care e mașina de predicție. Experiența perceptivă a voinței este o predicție perceptuală autosuficientă, un alt tip de halucinație controlată – sau, din nou, care *ne controlează*.

Mai este un motiv pentru care trăim acțiunile voluntare în felul în care le trăim, un motiv care clarifică și mai mult diferența dintre voință ca inferență perceptuală și voință ca magie dualistă. Trăirea voinței este necesară la fel de mult pentru ghidarea comportamentului *viitor*, ca pentru ghidarea comportamentului *curent*.

Scanned with OKEN Scanner

Așa cum am văzut, comportamentul voluntar este foarte flexibil. Abilitatea de a controla grade mari de libertate se traduce prin aceea că, dacă o anumită acțiune voluntară se sfârșește prost, data viitoare, într-o situație asemănătoare, aș putea încerca altceva. Dacă într-o luni o iau pe scurtătură ca să ajung la serviciu și întârzii, pentru că m-am rătăcit, atunci în ziua următoare aleg o rută mai lungă, dar pe care o știu. Experiențele voliționale semnalează diverse cazuri de comportament voluntar în așa fel încât să putem acorda atenție consecințelor și să ajustăm comportamentele viitoare pentru a ne atinge mai bine obiectivele.

Am menționat anterior că simțul liberului arbitru se referă mult la sentimentul că „am fi putut proceda altfel". Acest aspect contrafactual al experienței voliționale este foarte important pentru funcția sa orientate spre viitor. Sentimentul că aș fi putut proceda diferit *nu* înseamnă și că chiar aș fi putut proceda diferit. Mai degrabă, fenomenologia posibilităților alternative mă ajută într-o situație viitoare asemănătoare, dar nu identică, în care chiar aș putea proceda diferit. Dacă marți toate circumstanțele sunt aceleași ca luni, atunci marți nu pot proceda diferit față de luni. Numai că nu e niciodată cazul. Lumea fizică nu se duplică pe sine de la o zi la alta, nici măcar de la o milisecundă la alta. În final, circumstanțele creierului meu se vor fi schimbat, deoarece am trecut printr-o experiență volițională luni și am fost atent la consecințe. Asta este, în sine, suficient pentru a influența felul în care creierul meu poate controla multele mele grade de libertate atunci când se pregătește de serviciu marți.* Utilitatea sentimentului că „puteam proceda altfel" constă în aceea că, data viitoare, chiar ai putea.

Și cine este „tu"? „Tu" este ansamblul de convingeri, valori, scopuri, amintiri anterioare, relative la sinele tău, și

* Heraclit: „Niciun om nu intră în același râu de două ori, deoarece nu este același râu și nici omul nu este același".

Scanned with OKEN Scanner

cele mai bune estimări perceptive care, împreună, alcătuiesc experiența de *a fi tu*.

Experiențele voliționale în sine pot fi privite acum ca parte esențială a acestui întreg al sinelui personal – sunt o altă specie de halucinații controlate sau care ne controlează ale sinelui nostru. Împreună, capacitatea de a exersa și de a experimenta „liberul arbitru" sunt capacitatea de a realiza acțiuni, de a face alegeri – și de a gândi... gânduri – caracteristice ție.

Este liberul arbitru o iluzie, deci? Auzim adesea declarații înțelepte că așa ar fi. Renumitul psiholog Daniel Wegner a surprins ideea în cartea sa, *The Illusion of Conscious Will/Iluzia voinței conștiente*, care a rămas de căpătâi încă de la publicarea ei, acum aproape 20 de ani. Răspunsul corect la această întrebare este, desigur, „depinde".

Liberul arbitru ciudat cu siguranță nu e real. De fapt, liberul arbitru ciudat nu se califică nici măcar pentru categoria iluzoriu. Examinată îndeaproape, așa cum am văzut, fenomenologia voinței nu se referă atât la cauze imateriale, fără cauză, ci este o halucinație autosuficientă care ne controlează, referitoare la anumite tipuri de acțiuni – acele acțiuni care par să vină din interior. Privit astfel, liberul arbitru ciudat este o soluție incoerentă pentru o problemă care nu există.

Și, deși m-am concentrat, în acest capitol, pe exemple în care acțiunile voluntare sunt acompaniate de experiențe intense de voință, nu este întotdeauna cazul. Când cânt la pian sau fac o cană de ceai, în majoritatea timpului, aceste acțiuni voluntare se desfășoară cu un automatism și o fluență care subminează nu numai ideea naivă că *eu* aș putea fi cel care a produs aceste acțiuni, dar și pe cea mai puțin frecventată cum că astfel de acțiuni par să fie produse

Scanned with OKEN Scanner

de altceva. Atunci când oamenii descriu că sunt „în prezent" sau în „starea de flow" – imersați profund într-o activitate pe care au practicat-o multă vreme – fenomenologia voinței poate să lipsească cu desăvârșire. Mare parte din timp, acțiunile noastre voluntare și gândurile noastre – ei bine, „se întâmplă", pur și simplu. Referitor la liberul arbitru, nu e vorba doar despre faptul că felul în care par lucrurile nu este totuna cu felul în care și sunt. Cum *par* lucrurile merită o examinare mai detaliată, de asemenea.

Din altă perspectivă, liberul arbitru nu este deloc iluzoriu. Cât timp avem creiere relativ nedeteriorate și un fond relativ normal, fiecare dintre noi are posibilitatea de a executa și de a inhiba acțiunea voluntară, ca urmare a capacității creierului nostru de a ne controla multele grade de libertate. Acest tip de libertate este atât *față de*, cât și *de a*. Este o libertate *față de* cauzele imediate din corp și din afara lui, față de coerciția exercitată de autorități, hipnotizatori și înșelători, sau de promotorii din social media. Nu este, însă, libertate față de legile naturii sau de natura cauzală a universului. Este libertate *de* a acționa conform convingerilor, valorilor și țelurilor tale, de a face ceea ce îți dorești să faci și de a alege în virtutea a cine ești.

Realitatea acestui tip de liber arbitru este subliniată de faptul că nu poate fi considerat de la sine înțeles. Leziunile cerebrale, dotarea genetică sau socială mai puțin avantajoase pot submina capacitatea de a ne exersa comportamentul voluntar. Persoanele cu sindromul mâinii anarhice fac acțiuni voluntare pe care nu le percep ca fiind ale lor, în timp ce persoanele cu mutism akinetic sunt incapabile să facă vreo acțiune voluntară, oricare ar fi ea. O tumoră cerebrală situată într-un anume fel poate transforma un student la inginerie într-un ucigaș, așa cum s-a întâmplat în cazul lui Charles Whitman, „trăgătorul din Turnul Texas", sau să inducă unui profesor, până atunci ireproșabil,

Scanned with OKEN Scanner

o pedofilie galopantă – care a și dispărut după înlăturarea tumorii, ca să reapară când tumora a crescut la loc.

Dilemele etice și juridice formulate după astfel de cazuri sunt și ele cât se poate de reale. Charles Whitman nu a ales să aibă o tumoră cerebrală care să-i preseze amigdala adânc în creier, așa că ne putem întreba dacă este responsabil de acțiunile sale. La prima vedere, pare că nu, dar, pe măsură ce înțelegem mai mult despre baza cerebrală a voinței, nu devine totul un caz de „tumoră cerebrală până la capăt" pentru fiecare dintre noi?* Acest argument funcționează și în sens invers. Într-un interviu din 1929, Einstein a susținut că, deoarece nu credea în liber arbitru, nu își asuma niciun merit pentru nimic.

Este, de asemenea, o greșeală să afirmăm despre *experiența volițională* că este o iluzie. Aceste experiențe sunt cele mai bune estimări perceptive, la fel de reale precum alte tipuri de percepție conștientă, a lumii sau a sinelui. O intenție conștientă este la fel de reală ca experiența vederii culorilor. Niciuna nu corespunde direct unei anumite proprietăți definite a lumii – nu există „roșu adevărat" sau „albastru adevărat", așa cum nu există liber arbitru ciudat –, dar contribuie, ambele, într-o manieră deosebită la ghidarea comportamentului nostru și sunt, ambele, constrânse de convingeri anterioare și de date senzoriale. În timp ce experimentarea culorilor construiește trăsăturile lumii din jur, experimentarea voinței are acel conținut subversiv metafizic cum că „sinele" are influență cauzală asupra lumii. Proiectăm putere cauzală în experiențele

* Sistemele juridice vestice sunt fondate pe principiul că răspunderea penală necesită atât un „act vinovat" (*actus rea*), cât și o „minte vinovată" (*mens rea*). Atunci când capacitatea unei persoane de a-și exercita liberul arbitru – de a-și controla gradele de libertate – este afectată sau reprimată în vreun fel, se mai poate spune că au „mintea vinovată"? Unii, precum filosoful Bruce Waller, susțin că, atât timp cât nu noi decidem să avem creierul pe care îl avem, însuși conceptul de responsabilitate morală este incoerent. O altă perspectivă, de care sunt atras, este aceea că, odată ce depășim un anume prag al capacității de a ne controla gradele de libertate, putem fi considerați responsabili de acțiunile noastre.

Scanned with OKEN Scanner

noastre voliționale în același fel în care proiectăm roșul în percepția suprafețelor. Știind că această proiecție are loc – ca să-i dăm iar cuvântul lui Wittgenstein – ambele schimbă tot și lasă tot la fel cum era.

Experimentarea voinței nu este doar reală, este indispensabilă supraviețuirii noastre. Sunt inferențe perceptive autosuficiente care declanșează acțiuni voluntare. Fără aceste experiențe, nu vom fi capabili să navigăm prin mediile complexe în care noi, oamenii, prosperăm și nici să învățăm din acțiunile voluntare precedente pentru ca data viitoare să facem mai bine.

Briony Tallis credea că dacă ar putea identifica creasta valului, gata să se spargă, al acțiunii volitive, s-ar fi găsit pe sine. Sinele în chestiune este, bineînțeles, un sine uman și chiar pare să existe ceva tipic uman în capacitatea noastră de a face față mediilor complexe și schimbătoare prin comportament voluntar flexibil. Totuși, capacitatea de a ne exersa liberul arbitru ar putea să se manifeste pe grade, nu numai la noi, oamenii, ci și la animalele cu care împărțim lumea.

Iar dacă abilitatea de a ne exersa liberul arbitru se regăsește și la alte specii, ce se poate spune, atunci, despre extinderea conștienței în sine?

E timpul să privim dincolo de om.

Scanned with OKEN Scanner

IV

ALTELE

Scanned with OKEN Scanner

12
Dincolo de om

Începând de pe la mijlocul secolului al XVIII-lea, până în zorii secolului al XIX-lea, nu era neobișnuit, pentru curțile de justiție europene, ca animalele să răspundă penal pentru acțiunile lor. Porci, tauri, cai, elani, câini, chiar și delfini erau executați sau arși de vii. În aproximativ 200 de cazuri documentate de E. P. Evans în istoria sa, din 1906, a urmăririi penale la animale, porcii erau acuzați cel mai frecvent, probabil pentru că circulau liber prin orașele medievale. Crimele lor variau, de la mâncatul copiilor la al biscuiților sfințiți. Uneori, erau acuzați de complicitate – prin grohăiturile și pufniturile lor – la crimele altora; erau spânzurați adesea, uneori erau achitați.

Fig. 20: O scroafă și purceii ei sunt judecați pentru omorârea unui copil.
Fotografie din Arhiva de istorie universală, prin Getty Images.

Scanned with OKEN Scanner

Invaziile șoarecilor, lăcustelor, gărgărițelor și ale altor animale mici se rezolvau mult mai greu prin intermediul procedurilor legale. Într-un caz notoriu din secolul al XVI-lea, judecătorul francez Bartholomew Chassenée a exonerat niște șobolani, argumentând că nu se punea problema ca ei să apară la proces, din cauza pericolelor la care ar fi fost supuși prin prezența pisicilor pe tot parcursul drumului lor. În alte cazuri, incluzând diferite infestări cu gărgărițe, animalelor ofensatoare li se transmiteau ordine scrise de a părăsi o proprietate sau o cultură de orz, adesea într-o anumită zi și chiar la o anumită oră.

Oricât de bizare ar părea toate acestea pentru modul nostru de gândire de secol XXI, perspectiva medievală despre mintea animalelor a prefigurat recenta renaștere a interesului pentru conștiența animală și pentru ideea de a extinde noțiunea de „persoană" dincolo de om.* Ideea că animalele ar putea să înțeleagă și să se supună, în mod rezonabil, procedurilor ezoterice ale legilor ecleziale a fost și este aproape nebunească. Dar odată cu această idee vine și recunoașterea că animalele pot avea experiențe conștiente și ar putea avea minți capabile – în anumite circumstanțe – să ia decizii. Această recunoaștere a existenței minților conștiente dincolo de mintea umană este în contrast puternic cu versiunea carteziană a ideii de animal-mașină, în care animalele sunt lipsite de starea conștientă ce însoțește mințile raționale. Pentru medievali, animalele erau, cu siguranță, doar niște fiare. Dar nu animalele-robot ale dualismului cartezian, ci unele cu univers interior.

Astăzi ar fi ciudat, chiar pervers să susții că doar oamenii sunt conștienți. Dar ce am putea spune despre cât de mult se extinde cercul conștienței și despre cât de diferit ar putea fi universul interior al altor animale?

* Voi folosi cuvântul „animal" ca prescurtare pentru animalele „non-umane". Oamenii sunt, și ei, tot animale.

Scanned with OKEN Scanner

Primul lucru care trebuie spus este că nu putem decide dacă un animal este conștient după capacitatea – sau incapacitatea – de a ne comunica în vreun fel că este conștient. Lipsa unui limbaj nu este dovada lipsei conștienței. Nici absența așa-numitelor abilități cognitive „înalte" precum metacogniția – capacitatea de a reflecta asupra propriilor gânduri și percepții.

La animale, conștiența, acolo unde există, e diferită – în unele cazuri, foarte diferită – de a noastră. Deși experimentele pe animale pot arunca o lumină asupra mecanismelor conștienței la oameni, nu e înțelept să inferăm existența conștienței la animale doar pe baza unei similarități superficiale cu *Homo sapiens*. Procedând astfel, riscăm să deviem spre *antropomorfism* – atribuirea de calități umane non-umanului – și către *antropocentism* – tendința de a interpreta lumea doar în termeni de valori și experiențe umane. Antropomorfismul ne face să vedem conștiență acolo unde s-ar putea să nu fie – ca atunci când credem despre câinele nostru că înțelege ceea ce gândim. Antropocentrismul, pe de altă parte, ne face orbi la diversitatea minților animale, împiedicându-ne să recunoaștem conștiența care nu seamănă cu cea umană acolo unde s-ar putea să existe – o miopie exemplificată de perspectiva carteziană despre animalele-mașini.

Mai presus de orice, ar trebui să fim prudenți în asocierea prea strânsă a conștienței cu inteligența. Conștiența și inteligența nu sunt același lucru. Folosirea inteligenței ca indicator pentru conștiență duce la o serie de erori. Încalcă și antropomorfismul: constatăm că animalul X are o inteligență umană, animalul Y, nu, deci animalul X este conștient, iar Y, nu. Și se încurajează o lene metodologică, deoarece acceptarea capacităților „inteligente", precum limbajul și metacogniția – care sunt mai ușor de evaluat

Scanned with OKEN Scanner

decât conștiența în sine – ar fi suficientă pentru a infera conștiența.

Inteligența nu e lipsită de relevanță pentru conștiență. *Ceteris paribus*, inteligența oferă noi oportunități experienței conștiente. Poți fi trist sau dezamăgit fără să fie nevoie de prea multă competență cognitivă, dar, ca să simți regret – sau regret anticipat –, e nevoie de suficientă capacitate mentală pentru a lua în calcul rezultatele și direcțiile de acțiune alternative. După cum sugerează un studiu, chiar și șobolanii pot simți mai degrabă regret (în varianta rozătoarelor), decât dezamăgire atunci când lucrurile nu se petrec așa cum au sperat.*

Inferențele despre conștiența non-umană trebuie făcute cu multă prudență, ferindu-ne să impunem viziunea noastră antropocentrică. În același timp, nu prea avem opțiuni, în afara omului, în ce privește raportarea la o cantitate cunoscută; o bază solidă de la care să ne putem extinde. Până la urmă, știm că suntem conștienți și că deținem o înțelegere tot mai profundă a creierului și a mecanismelor implicate în conștiența umană și pe care le putem folosi ca bază pentru extrapolare.

Teoria animal-mașină dezvoltată în această carte susține despre conștiență că este mai bine conectată cu a fi *viu*, decât cu a fi *inteligent*. Asta se aplică în mod natural și la alte animale, nu numai la om. Din această perspectivă, conștiența poate fi mai extinsă decât pare, dacă am considera inteligența drept criteriu principal. Dar asta nu înseamnă că oriunde este viață, acolo e și conștiență.

Căutând conștiența dincolo de om este ca și cum ai păși de pe un mal înghețat pe lacul încremenit de lângă. Câte un pas, prudent, verificând cât de groasă e gheața de sub picioare.

* Într-un studiu din 2014, șobolanii trebuiau să aleagă între diferite opțiuni asociate cu diferite niveluri de răsplată. După ce au ales o variantă pentru care răsplata a fost mai mică decât se așteptau, șobolanii au devenit mai interesați de celelalte variante. Cercetătorii au interpretat acest comportament ca pe un semn de regret, deși e foarte greu de știut dacă șobolanii chiar simțeau – dacă simțeau – ceva.

Scanned with OKEN Scanner

Să începem cu mamiferele – grup ce include șobolani, lilieci, maimuțe, lamantini, lei, hipopotami și, firește, oameni. Eu cred că *toate mamiferele sunt conștiente*. Bineînțeles, nu știu sigur asta, dar sunt destul de convins. Această afirmație nu se bazează pe similaritatea superficială cu omul, ci pe un mecanism comun. Dacă omiți dimensiunea generală a creierului – care se corelează mai degrabă cu mărimea corpului decât cu orice altceva – creierul mamiferelor este surprinzător de similar la diferite specii.

În 2005, Bernard Baars, specialist în științe cognitive, eu și David Edelman – fiul lui Gerald Edelman, expert în cogniția animalelor – am făcut o listă a proprietăților conștienței umane care ar putea fi testate cu ușurință la alte animale. Am înșirat 17 astfel de proprietăți. Era un număr arbitrar, în anumite privințe, dar a demonstrat că formularea de întrebări testabile despre conștiența animalelor este rezonabilă.

Primele proprietăți la care ne-am gândit aveau de-a face cu trăsăturile anatomice ale creierului. În ce privește cablarea cerebrală, trăsăturile neuro-anatomice strâns asociate cu conștiența umană se regăsesc la toate speciile de mamifere. Există un cortex cu 6 straturi, un talamus interconectat intim cu cortexul, un trunchi cerebral profund și o serie de alte aceleași caracteristici – inclusiv sistemele de neurotransmițători – care sunt implicate în mod consecvent, la om, în desfășurarea, de la o clipă la alta, a experienței de a fi conștient.

Există trăsături comune și în cazul activității cerebrale. Printre cele mai uimitoare, sunt schimbările care intervin în dinamica cerebrală atunci când animalele adorm și se trezesc – dinamică ce susține *nivelul* de conștiență. În stările de veghe normale, toate mamiferele au activitate electrică cerebrală rapidă, neregulată, de amplitudine joasă.

Scanned with OKEN Scanner

Atunci când adorm, dinamica se schimbă către ritmuri mai regulate, cu amplitudine mare. Aceste tipare și schimbări se aseamănă mult cu ceea ce observăm la oameni în perioadele de veghe și de somn. Anestezia generală are și ea efecte similare la diferite specii de mamifere – o pauză generalizată în comunicarea dintre zonele cerebrale, care însoțește lipsa totală de răspuns comportamental.

Există, evident, și numeroase diferențe, mai ales în tiparele somnului. Focile și delfinii dorm doar cu câte o jumătate de creier, koala dorm circa 22 de ore zilnic, girafele, mai puțin de 4, iar puii orcilor nu dorm deloc în prima lună după naștere. Aproape toate mamiferele au somn REM (rapid eye movement = mișcări rapide ale ochilor) – focile, doar atunci când dorm pe țărm, iar delfinii se pare că nu au somn REM deloc.

În afară de nivelul de conștiență, vor fi și diferențe substanțiale în ce privește *conținutul* de conștiență la diferite specii de mamifere. Mare parte a acestor variații pot fi atribuite diferențelor de la nivelul tipurilor dominante de percepție. Șoarecii se bazează pe mustăți, liliecii, pe sonar, iar Heterocephalus glaber (șobolanul-cârtiță golaș), pe simțul extrem de ascuțit al mirosului – mai ales atunci când se întâlnesc cu alții ca ei. Aceste diferențe în dominanța perceptivă înseamnă să fiecare animal populează un univers interior distinctiv.*

Și mai surprinzătoare sunt diferențele legate de trăirile *sinelui*. La oameni, un marker important al dezvoltării conștienței de sine de nivel înalt, comparabil cu cel legat de identitatea personală, este abilitatea de a se recunoaște în oglindă. Această capacitate de „autorecunoaștere în oglindă" tinde să se dezvolte, la om, undeva între vârsta de 18 și 24 de luni. Asta nu înseamnă că bebelușii nu sunt conștienți – ci doar că acea conștiență de sine, ca individ

* Lumea, așa cum e trăită de un animal, este adesea numită *Umwelt* al acelui animal – termen introdus de etologul Jakob von Uexkull.

Scanned with OKEN Scanner

separat de ceilalți, se poate să nu se fi format înainte de vârsta respectivă.

La animale, capacitatea de autorecunoaștere a fost investigată pe larg cu ajutorul unui test dezvoltat de psihologul Gordon Gallup Jr, în anii 70. În versiunea clasică a testului de autorecunoaștere, un animal este anesteziat și, apoi, marcat cu o vopsea sau cu o bulină, pe o parte a corpului pe care nu și-o poate vedea, îndeobște. Când efectul anesteziei se estompează, animalului i se dă ocazia să interacționeze cu o oglindă, în așa fel încât să poată vedea marcajul. Dacă, după ce interacționează cu oglinda, animalul caută marcajul pe corpul său în loc să investigheze imaginea din oglindă, atunci a trecut testul. Criteriul se bazează pe raționamentul conform căruia ființa și-a dat seama că imaginea din oglindă este a ei, nu a altui animal.

Cine trece testul oglinzii? Dintre mamifere, maimuțele antropoide, câțiva delfini, orcile și un singur elefant eurasiatic. O listă întreagă de mamifere, printre care urșii panda, câinii, diferite maimuțe, l-au picat – cel puțin până acum. Având în vedere cât de intuitivă este pentru noi, oamenii, autorecunoașterea în oglindă, și cât de competente cognitiv par să fie multe dintre aceste mamifere, care nu se recunosc, însă, pe ele, această listă a celor care au trecut testul este surprinzător de scurtă. Nu există nicio dovadă convingătoare că vreun non-mamifer trece testul oglinzii, deși rechinii din genul Manta și coțofenele s-ar putea califica, precum și, mai nou și cu oarecare controverse, peștii din familia Labridaelor.

Animalele ar putea pica testul oglinzii din mai multe motive decât lipsa abilităților de a se recunoaște pe ele însele. Printre acestea, s-ar putea număra faptul că nu le plac oglinzile, că nu înțeleg cum funcționează sau chiar evitarea contactului vizual. Recunoscând aceste motive, cercetătorii pot dezvolta noi versiuni ale testului, care să

Scanned with OKEN Scanner

fie mai bine adaptate la diferite universuri interioare – diferite lumi perceptive. De exemplu, autorecunoașterea la câini poate fi acum testată cu „oglinzi olfactive" – deși nici acestea nu sunt foarte bune (cogniția la câini este numită, încântător, „dogniție" – de la dog = câine, n. trad.). Este posibil ca, odată cu dezvoltarea ingenuității experimentale, specii care se situau „sub linie" să treacă deasupra ei în ce privește trecerea testului de autoconștientizare cu ajutorul oglinzii. Chiar și așa, diversitatea diferitelor teste ale oglinzii – și incapacitatea multor animale de a trece teste mult mai bine adaptate speciei – sprijină ideea diferențelor uriașe între mamifere în ceea ce privește experiența de „a fi ele însele".

Aceste diferențe mă surprind de-a binelea mai ales în cazul maimuțelor. Deși cimpanzeii și maimuțele antropoide ne sunt cele mai aproape din punct de vedere evolutiv, maimuțele nu sunt cu mult mai departe și au fost folosite de multă vreme în experimentele neuroștiințifice ca „model primată" pentru om, mai ales în studiul vederii. În unele studii, maimuțele au fost dresate să „raporteze" – prin apăsarea unei pedale, de exemplu – pentru a „spune" ceva. Aceste experimente pot fi comparate direct cu studiile pe oameni, în care subiecții spun ceea ce simt sau nu, oferind un echivalent primat al unei metode esențiale în cercetarea conștienței.

Din cauza asemănării foarte mari cu omul, sunt convins că maimuțele au un fel de sine conștient. Dacă petreci suficient de mult timp cu maimuțele, impresia că te afli printre alte entități conștiente – alți *sine* conștienți – este covârșitoare.

Am trecut prin asta în iulie 2017, când am petrecut o zi la Cayo Santiago, o mică insulă foarte aproape de

Scanned with OKEN Scanner

coasta estică a Puerto Rico, în Caraibi. Cayo Santiago este cunoscută și ca „insula maimuței", deoarece singurii ei locuitori sunt macaci rhesus – mai mult de o mie. Populația a fost mutată aici în 1938 din Kolkata de către un zoolog american excentric, Clarence Ray Carpenter, care obosise să le tot urmărească prin toată India. M-am plimbat prin Cayo Santiago în acea zi de vară fierbinte, în compania psihologului de la Yale, Laurie Santos (și o echipă de filmare), printre zeci de maimuțe care își vedeau de treabă, prudente, dar și tolerante cu noi, oamenii greoi. Două maimuțe urcau, pe rând, într-un copac, de unde se aruncau într-un iaz de dedesubt și mi-a fost clar că singurul motiv pentru care făceau asta era pura plăcere spontană. Se amuzau.*

La fel de convingătoare sunt și înregistrările cu maimuțe capucin care reacționează la nedreptatea deliberată. În una dintre acestea, popularizată de primatologul Frans de Waal, două maimuțe, așezate în cuști alăturate, sunt răsplătite, una după alta, pentru înmânarea unei pietricele experimentatorului. Maimuța nr. 1 înmânează piatra prin plasa cuștii și primește o bucățică de castravete, pe care o consumă cu plăcere. Maimuța nr. 2 face același lucru, dar primește un strugure, nu un castravete – ceva mult mai gustos. Maimuța nr. 2 consumă strugurele, în timp ce maimuța nr. 1 se uită. Atunci când Maimuța nr. 1 repetă gestul și primește iar castravete, îl privește, apoi îl aruncă înapoi experimentatorului, zgâlțâind cușca în semn de ceea ce pare a fi indignare.

Amuzamentul și tantrumurile sunt „automate de persuasiune"/„intuition pumps" foarte puternice (concept formulat de Daniel Dennet, care se referă la un experiment

* La scurt timp după vizita noastră, Cayo Santiago – împreună cu mare parte a Puerto Rico – a fost devastată de uraganul Maria. Din fericire, majoritatea maimuțelor a supraviețuit. Mare parte a infrastructurii de cercetare, însă, a fost distrusă. Înregistrările făcute atunci apar în documentarul din 2018 intitulat The Most Unknown/Cel mai necunoscut – (www.themostunknown.com).

Scanned with OKEN Scanner

de gândire făcut să dirijeze audiența către o anumită concluzie prin intuiție – n. trad.). Aceste comportamente sunt atât de caracteristice, încât este aproape imposibil să le interpretăm ca fiind altceva decât manifestarea, către exterior, a unor stări interioare similare celor umane. Atunci când privim o maimuță comportându-se astfel, deducem prezența nu numai a unei alte ființe conștiente, ci a unei ființe conștiente *ca noi*. Așa cum am precizat, însă, maimuțele au picat mereu testul oglinzii. În timp ce e clar că maimuțele sunt conștiente, iar eu sunt convins că au un soi de sine, ele nu sunt niște omuleți blănoși.

Modelarea intuițiilor noastre de către antropomorfism și antropocentrism devine și mai evidentă atunci când ne uităm dincolo de mamifere. Mai ales când privim mult mai departe, către cele mai îndepărtate rude în linie evolutivă.

În vara lui 2009, David Edelman și cu mine am petrecut o săptămână cu circa o duzină de *Octopus vulgaris* – caracatița comună, o specie de cefalopod.* Îl vizitam pe biologul Graziano Fiorito, expert de marcă în cogniția și neurobiologia cefalopodelor. Deși a trecut mai bine de un deceniu, acea săptămână a rămas una dintre cele mai memorabile din toți anii mei ca cercetător.

Laboratorul cu caracatițe al lui Fiorito – parte a renumitului institut italian de cercetare Stazione Zoologica – se află în subsolul umed de sub acvariul public din inima orașului Napoli, un refugiu răcoros de la căldura verii de deasupra. Mi-am petrecut acea săptămână mai ales alături de aceste creaturi fascinante,

* Cefalopodele includ caracatițe, calmari și sepii, precum și ființe simple ca nautilul, printre circa 800 de specii pe care de dispariție. Termenul de „cefalopod" se traduce literar prin „cap cu picioare", traducere destul de nefericită pentru caracatițe, care au, mai degrabă, apendici asemănători cu brațele, decât picioare atașate la cap.

Scanned with OKEN Scanner

observând cum își schimbă forma, culoarea și textura și fiind atent la ce anume sunt ele atente. Într-o zi, pe când încercam să potrivesc aspectul în continuă schimbare al unei anumite caracatițe cu desenele lui Fiorito din *A Catalogue of Body Patterning in Cephalopoda* (Un catalog al tiparelor corporale la Cefalopode), am auzit un plescăit greoi și o alunecare. Lăsasem întredeschis capacul unui acvariu, iar ființa evadase. Și astăzi sunt convins că m-a păcălit, inducându-mi o falsă senzație de siguranță, și așteptând să mă întorc cu spatele suficient de mult.

În paralel cu ravagiile pe care le produceam, David pregătea un experiment referitor la percepția vizuală și învățare. Cobora obiecte de forme diferite în acvariul unei caracatițe, unele dintre ele însoțite de un crab gustos. Ideea era de a verifica dacă acea caracatiță poate învăța să asocieze anumite obiecte cu o răsplată. Nu-mi amintesc cum anume a evoluat studiul, dar îmi amintesc foarte clar un episod.

Laboratorul lui Fiorito avea două rânduri de acvarii, fiecare cu o caracatiță, aliniate de o parte și de alta a unei alei centrale. (Caracatițele nu prea sunt sociale, putând fi chiar canibale.) În acea anume zi, David alesese un acvariu de pe partea stângă a aleii, situat cam la jumătatea rândului. Când am intrat ca să văd ce se întâmplă, am rămas uimit să constat că toate caracatițele din acvariile de pe rândul opus erau lipite de geamuri, privind intens la David în timp de acesta cobora repetat obiectele în acvariul ales de el. Părea că încearcă, cu toatele, să-și dea seama de ce anume se întâmplă, fără niciun alt motiv decât pura curiozitate.

Traiul printre caracatițe, chiar și pentru o perioadă scurtă, mi-a lăsat impresia că sunt creaturi inteligente și conștiente, foarte diferite una de cealaltă – și, cu siguranță, foarte diferite de noi. Evident, este o impresie subiectivă, marcată de antropomorfism și antropocentrism, și dispusă să considere inteligența drept semn al capacității de a simți.

Scanned with OKEN Scanner

Caracatița este remarcabilă ea însăși, în mod obiectiv, iar observarea lor ne poate ajuta să înțelegem cât de diferită poate fi o conștiență non-umană.

Cel mai recent strămoș comun al omului și caracatiței a trăit acum cca 600 de milioane de ani. Se știe foarte puțin despre această creatură. Poate că era un soi de vierme plat. Indiferent de cum anume arăta, trebuie că era un animal foarte simplu. Mințile caracatițelor nu sunt interpretări acvatice ale minților noastre sau ale altor specii cu coloană vertebrală, foste sau prezente. Mintea unei caracatițe este un experiment evolutiv independent, la fel de aproape de mintea unui extraterestru pe cât e posibil să întâlnim unul pe această planetă.

Așa cum spune filosoful scufundător Peter Godfrey-Smith, „dacă vrem să înțelegem *altfel de* minți, ale cefalopodelor sunt cele mai altfel dintre toate."

Corpul unei caracatițe este suficient de remarcabil. Caracatița comună, *O. vulgaris*, are opt apendici asemănători unor brațe, trei inimi care pompează sânge albastru, un mecanism de apărare pe bază de cerneală și propulsie cu jet, extrem de sofisticată. O caracatiță își poate schimba dimensiunea, forma, textura și culoarea după bunul plac, chiar pe toate în același timp, dacă e nevoie. Este un animal lichid: în afară de un cioc osos situat central, restul corpului este moale, putându-se strecura prin fante extrem de fine – după cum descoperisem la Stazione Zoologica.

Aceste corpuri extraordinare au un sistem nervos extrem de sofisticat. *Octopus vulgaris* are aproape jumătate de milion de neuroni, de circa șase ori mai mult decât un șoarece. Spre deosebire de neuronii mamiferelor, majoritatea neuronilor la caracatițe – aproape trei cincimi – se află mai degrabă în brațe decât în creierul central, creier dotat cu circa 40 de lobi anatomic distincți. O altă trăsătură neobișnuită e că neuronii caracatițelor nu au mielină – materialul izolator care, la mamifere, permite dezvoltarea

Scanned with OKEN Scanner

și funcționarea conexiunilor neuronale pe distanțe lungi. Sistemul nervos al caracatiței este, astfel, mai distribuit și mai puțin integrat decât sistemul nervos la mamifere de dimensiuni și complexitate similare. Conștiența caracatiței – presupunând că există așa ceva – ar putea fi, deci, mai distribuită și mai puțin integrată, poate fără un „centru" singular.

Caracatițele fac altfel lucrurile chiar și la nivel de gene. La majoritatea organismelor, informația genetică din ADN este transcrisă direct în secvențe scurte de ARN (acid ribonucleic) care sunt, apoi, folosite pentru sinteza de proteine – caii de povară moleculari ai vieții. Acesta este un principiu bine stabilit, de nivel de manual, al biologiei moleculare. În 2017, acest principiu avea să fie contrazis de descoperirea faptului că secvențele de ARN de la caracatițe – și de la câteva alte cefalopode – pot suferi editări semnificative înainte să fie traduse în proteine. E ca și cum caracatița își poate rescrie părți din propriul genom pe parcurs. (Editarea ARN a fost identificată anterior și la alte specii, dar în acele cazuri avea rol relativ minor.) Mai mult, în cazul caracatiței, mare parte a editării ARN pare să fie legată de sistemul nervos. Unii cercetători au sugerat că această capacitate prolifică de rescriere a genomului ar putea explica, măcar parțial, capacitățile cognitive remarcabile ale caracatițelor.

Și capacitățile lor cognitive sunt, cu siguranță, impresionante. Pot găsi obiecte ascunse – de obicei, crabi gustoși – în cuburi de plexiglas puse unele în altele, se descurcă în labirinturi complexe, încearcă diferite acțiuni pentru a rezolva o anumită problemă și – așa cum a arătat Fiorito la Stazione – învață doar din simpla observare a celorlalte caracatițe. Rapoartele anecdotice ale comportamentelor caracatițelor în sălbăticie sunt și mai uimitoare. În unul dintre exemplele extraordinare, difuzate în seria BBC *Blue Planet II*, apare o caracatiță, surprinsă în câmp deschis,

Scanned with OKEN Scanner

care se acoperă cu scoici și detritus de pe fundul oceanului pentru a se ascunde de rechinul prădător.

Aceste caracteristici ale inteligenței cefalopodelor sunt, cu siguranță, dovezi convingătoare ale existenței unei minți în funcțiune. Dar ce fel de minte? Am spus deja că nu ar trebui să considerăm inteligența ca reper pentru conștiență. Deci ce anume se poate spune despre cum este să fii o caracatiță? Pentru a găsi răspunsul, trebuie să conectăm comportamentul cu percepția caracatiței.

Camuflajul este, probabil, cea mai de-pe-altă-lume dintre caracteristicile cefalopodelor. Fără carapace solide pentru protecție, supraviețuirea lor depinde adesea de abilitatea de a se confunda cu fundalul. Pot lua atât de bine culorile, formele și texturile mediului lor înconjurător, încât tu sau eu sau alți potențiali prădători nu le putem vedea, nici măcar de la un metru.

Caracatițele iau culoarea împrejurimilor prin intermediul unui sistem extrem de precis de *cromatofori*. Aceștia sunt săculeți elastici, distribuiți pe toată suprafața epidermei, și care produc colorarea acesteia în roșu, galben sau brun, prin comandă neuronală ce provine de la lobul dedicat cromatoforilor, din creier. Cum anume funcționează acest mecanism nu se știe pe deplin. Motivul stă și în faptul că aceste animale devin invizibile nu numai pentru alte caracatițe, dar și pentru prădători care văd lumea în felul lor caracteristic. Sistemul lor de camuflaj codifică, probabil, cunoaștere despre capacitățile vizuale ale acelor prădători.

Și mai surprinzător este că acest comportament se manifestă deși caracatițele nu văd culorile. Celulele sensibile la lumină din ochiul uman răspund la trei lungimi de undă diferite ale luminii, creând, din amestecul lor, un univers de culoare. Celulele din ochiul de caracatiță, însă, conțin doar

Scanned with OKEN Scanner

un singur foto-pigment. Caracatița poate simți direcția polarizării luminii – la fel ca tine sau ca mine, când purtăm ochelari pentru polarizarea luminii – dar nu poate extrage culori din combinarea lungimilor de undă. Nici celulele sensibile la lumină din epiderma lor nu percep culorile: caracatițele pot „vedea" atât cu epiderma, cât și cu ochii. Mai mult, controlul cromatoforilor la caracatițe pare să fie „în buclă deschisă", ceea ce înseamnă că neuronii din lobul cromatoforilor nu generează nicio copie internă evidentă a semnalelor trimise către cromatoforii din epidermă. Creierul central pare să nici nu știe ce face epiderma.

Este greu să-ți dai seama ce anume înseamnă toate acestea pentru cum își experimentează o caracatiță lumea, precum și corpul aflat în acea lume. Propria ei epidermă își schimbă culoarea în feluri în care ea nu le poate vedea și care nici nu au legătură cu creierul ei. Iar parte a acestor adaptări pot avea loc doar prin control local, în care un braț simte mediul înconjurător apropiat și își schimbă aspectul fără ca creierul central să fie implicat în vreun fel. Presupunerea tipic umană potrivit căreia putem vedea și simți ce se întâmplă corpurilor noastre pur și simplu nu se aplică. Nu e surprinzător, deci, că testul oglinzii nu e trecut de către caracatițe.

La fel ca mamiferele și ca alte vertebrate, caracatițele au cele cinci simțuri fundamentale – văz, gust, miros, atingere și chiar și auz, acesta din urmă nu foarte bun. În mod bizar, caracatițele pot simți gustul nu numai cu gura centrală, ci și cu ventuzele, ceea ce indică descentralizarea remarcabilă a minții la aceste creaturi.

Ideea de conștiență descentralizată devine provocatoare atunci când ne referim la simțul de proprietate asupra propriului corp. Așa cum am văzut în capitolul 8, la oameni, acest aspect al sinelui conștient poate fi alterat surprinzător de ușor, păcălind creierul să-și schimbe cea mai bună estimare bayesiană despre ce anume este și ce

Scanned with OKEN Scanner

nu este parte a corpului. Controlul corpului nostru este suficient de dificil pentru noi, oamenii, cu ale noastre patru membre și articulațiile lor. Pentru o caracatiță, care are opt brațe extrem de flexibile, care se desfășoară și se înfășoară la loc în toate direcțiile, provocarea este extraordinară. Și, așa cum senzația este devotată parțial acestor brațe, tot așa și controlul. Brațele caracatiței se comportă precum niște animale semi-autonome: un braț retezat poate executa secvențe de acțiune complexă, precum apucarea hranei, la ceva timp după ce a fost separat de corp.

Împreună, aceste grade de libertate și controlul descentralizat reprezintă o sarcină complicată pentru un creier central, care încearcă să mențină o percepție unică și unificată a ceea ce este și ce nu este parte a corpului. Motiv pentru care caracatițelor s-ar putea să nu le pese. Oricât de ciudat ar putea să sune, ceea ce pare să fie o caracatiță ar putea să nu includă o experiență de proprietate asupra propriului corp în niciunul din sensurile cunoscute la oameni și la alte animale.

Asta nu înseamnă că ele nu fac diferența între „sine" și „alții". E clar că o fac – și trebuie să o facă. Pentru început, trebuie să evite să se încurce în ele însele. Ventuzele de pe brațele unei caracatițe se lipesc reflex de orice obiect în trecere și, totuși, nu se atașează de celelalte brațe ale aceleiași caracatițe și nici de zona ei centrală. Asta demonstrează că aceste ființe pot discrimina, într-un fel, ce anume e corpul lor și ce nu.

Se pare că această abilitate depinde de un sistem simplu, dar eficient, de autorecunoaștere pe bază de gust. Caracatițele secretă o anumită substanță chimică prin tegument. Această substanță servește ca semnal, care poate fi detectat de ventuze și, astfel, nu se lipesc automat. În acest fel, o caracatiță poate ști *ce* anume e parte a ei și ce nu e, chiar dacă nu știe neapărat *unde* se află corpul ei în spațiu. Această descoperire a fost făcută după o serie de

Scanned with OKEN Scanner

experimente mai degrabă macabre, în care cercetătorii au oferit unor brațe detașate alte brațe detașate, fie cu tegument, fie fără. Ventuzele s-au fixat automat de brațele fără tegument, nu și de cele intacte.*

Ce anume înseamnă asta pentru experiența întrupării la caracatiță este greu de imaginat pentru noi, mamiferele. Caracatița ca întreg ar putea avea doar o percepție vagă a ce este și unde se află corpul ei, deși experiența acestei percepții s-ar putea să nu fie vagă. Și ar putea fi chiar ceva despre cum e un braț de caracatiță.

Caracatițele ne pun la grea încercare impresia despre cât de diferită ar putea fi conștiența la animale comparativ cu omul. Trecând direct de la maimuțe la cefalopode, am omis, însă, o menajerie enormă. Departe de țărmurile sigure ale conștienței mamiferelor se află un vast potențial de conștiență animală, de la papagal până la parameciul unicelular. Contemplând această întindere, să revenim la întrebările fundamentale despre care animale ar putea să aibă orice formă de experiență conștientă – acele animale pentru care „lumina e aprinsă", chiar dacă acea lumină e doar o licărire.

Păsările reprezintă un caz foarte bun pentru subiectul simțirii. Creierul de pasăre, deși semnificativ diferit de cel de mamifer, are o organizare similară cortexului și talamusului de mamifer. Multe specii de păsări sunt deosebit de inteligente. Papagalii pot număra, cacadu pot dansa, gaițele depozitează hrana în funcție de viitoarele lor nevoi. În timp ce aceste exemple de inteligență sugerează că unele păsări s-ar putea bucura de stări complicate de

* Aceste experimente nu sunt chiar atât de oribile cum sună. Caracatițele nu par să observe când li se amputează un braț, în scurt timp le crește altul la loc. Bineînțeles, asta nu înseamnă că ar trebui să faci astfel de experimente dacă nu ai motiv suficient de bun.

Scanned with OKEN Scanner

conștiență, să ne amintim că inteligența nu este un test care confirmă conștiența. Păsările care nu ascund hrană, care nu vorbesc și nu dansează pot avea și ele stări de conștiență.

Pe măsură ce avansăm, dovezile devin mai rare și mai sumare, iar inferențele despre conștiență, mai provizorii. În loc să clădim aceste inferențe pe similaritatea cu creierul și cu comportamentul mamiferelor, o strategie mai bună ar fi să adoptăm perspectiva animal-mașină – a mea, nu versiunea carteziană –, care urmărește originea și funcția percepției conștiente până la reglarea fiziologică și la păstrarea integrității organismului. Ceea ce sugerează că putem căuta dovada conștienței în modul în care animalele răspund la evenimente despre care se presupune că sunt dureroase.

Această strategie nu este doar sensibilă din punct de vedere științific, ci și motivată de imperative etice. Deciziile referitoare la bunăstarea animalelor nu ar trebui să se bazeze pe asemănarea cu oamenii și nici pe depășirea vreunui prag arbitrar de competență cognitivă, ci pe susceptibilitatea la durere și suferință. Și, deși există nenumărate moduri în care creaturile vii pot suferi, cele mai răspândite implică, probabil, provocări elementare la adresa integrității lor fiziologice.

Atunci când au fost căutate, s-au găsit nenumărate dovezi pentru răspunsurile adaptative la evenimente dureroase în rândul speciilor de animale. Majoritatea vertebratelor (animale cu coloană vertebrală) vor indica zona afectată a corpului. Chiar și minusculii pești zebră vor alege să „plătească" pentru a scăpa de durere, trecând din mediul lor natural într-un acvariu steril și puternic luminat, dar plin de analgezic. Dacă asta implică faptul că peștii sunt conștienți – și există *multe* tipuri de pești – e neclar, dar, cu siguranță, sugestiv.

Dar insectele? Furnicile nu șchioapătă atunci când au un picior rupt. Exoscheletul lor dur ar putea să fie mai

Scanned with OKEN Scanner

puțin susceptibil la durere, iar creierul insectelor are forme de neurotransmițători din categoria opioidelor, asociați, de obicei, cu ușurarea durerii la alte animale. Un studiu recent a arătat că musculița de oțet, *Drosophila melanogaster*, manifestă o hipersensibilitate post-traumă la stimuli anterior nedureroși într-un fel asemănător „durerii cronice" de la oameni. Și, remarcabil, medicamentele anestezice par să aibă efect la *toate* animalele, de la organismele unicelulare până la primatele cele mai evoluate.

Toate acestea sunt sugestive, niciuna, însă, concludentă.

La un moment dat, devine dificil să spui ceva substanțial. Părerea mea – și nu e decât o părere – este că există animale care nu fac parte deloc din cercul conștienței. Unul dintre motivele pentru care cred asta este acela că, chiar și la mamifere, cu creierul nostru complex și sistemele perceptive fin acordate și orientate spre păstrarea integrității fiziologice, inconștiența rămâne relativ ușor de atins. Experiența conștientă este esențială pentru viața noastră, dar asta nu înseamnă că baza ei biologică ne este accesibilă. Până să ajungem la viermele nematod, cu ai săi doar 302 de neuroni, mi se pare dificil să atribui vreo stare conștientă semnificativă – iar parameciul unicelular nu îndeplinește criteriile.

Lăsând la o parte incertitudinile inevitabile, studiul conștienței la animale vine cu două beneficii esențiale. Primul este recunoașterea faptului că modul în care omul trăiește lumea și pe sine nu este unicul. Populăm o mică zonă dintr-un spațiu imens de posibile minți conștiente, iar investigarea științifică a acestui spațiu până acum înseamnă puțin mai mult decât câteva pâlpâiri în beznă. Al doilea este o smerenie regăsită. Privind la uriașa diversitate a vieții pe Terra, putem prețui mai mult – și

Scanned with OKEN Scanner

primi mai puțin de-a gata – bogăția experienței subiective în toată varietatea și individualitatea sa, în noi și în alte animale. Am putea găsi, de asemenea, o motivație reînnoită de a minimiza suferința, oriunde și oricum ar apărea.

Am început acest capitol argumentând despre conștiență și inteligență că nu sunt același lucru și, despre conștiență, că are mai mult de a face cu a fi viu decât cu a fi deștept. Vreau să-l termin cu o afirmație și mai puternică. Conștiența nu numai că există fără prea multă inteligență – nu e nevoie să fii deștept ca să suferi –, dar și inteligența poate exista fără conștiență.

Posibilitatea de a fi deștept fără să suferi ne aduce la etapa finală a călătoriei noastre prin știința conștienței. E timpul să vorbim despre inteligența artificială și despre posibilitatea ca mașinile conștiente să existe.

Scanned with OKEN Scanner

13
Minți-mașină

În Praga sfârșitului de secol XVI, rabinul Judah Loew ben Bezalel a luat lut de pe malul râului Vltava și a modelat o siluetă umană – un golem. Acest golem – numit Josef, sau Yoselle – avea ca sarcină să apere comunitatea rabinului de pogromurile anti-semite și se pare că a fost foarte eficient în acest sens. Odată activat printr-o incantație magică, golemul Josef se putea mișca, era conștient și se supunea. Dar cu Josef s-a întâmplat ceva teribil, el devenind, din supusul obedient, un monstru violent. În cele din urmă, rabinul și-a rupt vraja, iar golemul a căzut, în bucăți, pe podeaua sinagogii. Unii spun că rămășițele lui zac și acum, ascunse pe undeva prin Praga, poate într-un cimitir, poate într-un pod, poate așteptând, răbdător, să fie reactivat.

Golemul rabinului Loew ne amintește de orgoliul cu care proiectăm creaturi inteligente, simțitoare – după chipul și asemănarea noastră, sau din mintea lui Dumnezeu. Rareori se sfârșește cu bine. De la monstrul din *Frankenstein* al lui Mary Sheely, la Ava din *Ex Machina* a lui Alex Garland, trecând prin roboții eponimi ai lui Karel Čapek, *Terminator*-ul lui James Cameron, replicile din *Blade Runner* al lui Ridley Scott și HAL al lui Stanley Kubrick, aceste creații au ajuns să se întoarcă împotriva celor care i-au creat, lăsând în urma lor distrugere, tristețe și confuzie filosofică.

În ultimul deceniu, aproximativ, avansul AI relansează întrebările despre conștiența mașinilor. AI este peste tot acum, în telefoane, frigidere, mașini, fiind activat de algoritmi de rețea neuronală inspirați din arhitectura creierului.

Scanned with OKEN Scanner

Ne îngrijorăm pe bună dreptate de impactul acestei noi tehnologii. Ne va lua locurile de muncă? Va distruge structura socială umană? Ne va distruge, în cele din urmă, pe toți – fie pentru că vrea asta, fie din cauza lipsei unei previziuni a programării și care duce la transformarea tuturor resurselor Terrei într-un munte de agrafe de birou? Sub toate aceste îngrijorări, mai ales a celor existențiale și apocaliptice, se află presupunerea că AI va deveni – într-unul dintre momentele de dezvoltare accelerată – conștientă. Acesta este mitul golemului din siliciu.

Ce anume face ca o mașină să fie conștientă? Care ar fi implicațiile? Și cum vom putea face diferența dintre o mașină conștientă și al ei zombi echivalent?

De ce am crede că o mașină – o inteligență artificială – ar putea deveni conștientă? Așa cum am menționat deja, e destul de obișnuit – nu și universal – să credem despre conștiență că va apărea în mod natural odată ce mașinile trec un anume prag, deocamdată necunoscut, al inteligenței. Ce anume generează această impresie? Cred că două sunt presupunerile responsabile, niciuna justificabilă. Prima presupunere este despre condiția *necesară* pentru ca orice să fie conștient. A doua presupunere este despre condiția *suficientă* pentru ca ceva să fie conștient.

Prima presupunere – condiția necesară – este *funcționalismul*. Conform funcționalismului, conștiența nu depinde de părțile din care e făcut un sistem, fie wetware/organic, fie hardware/anorganic, fie porți logice neuronale, fie din siliciu – sau lut de pe malul Vltavei. Funcționalismul afirmă că ceea ce contează pentru conștiență este ceea ce *face* un sistem. Dacă un sistem transformă corect input-urile în output-uri, atunci există conștiență. Așa cum am explicat în capitolul I, aici sunt două afirmații separate. Prima

Scanned with OKEN Scanner

se referă la independența de orice substrat sau material anume, a doua, la suficiența relației input-output. În cea mai mare parte a timpului merg împreună, dar, uneori, se despart.

Funcționalismul este o perspectivă populară printre filosofii minții și e, adesea, acceptat ca poziție implicită și de către non-filosofi. Asta nu înseamnă însă că e și corect. Pentru mine, nu există argumente decisive nici pro, nici contra ideii conform căreia conștiența este independentă de substrat sau că ar fi doar o chestiune de relații input-output, de „procesare a informației". Atitudinea mea față de funcționalism este una de suspiciune agnostică.

Pentru ca un computer cu inteligență artificială să devină conștient, funcționalismul trebuie să fie adevărat. Aceasta este condiția necesară. Funcționalismul ca adevăr nu este, în sine, suficient pentru conștiență: procesarea informației în sine nu este suficientă pentru conștiență. A doua presupunere se traduce prin aceea că tipul de procesare a informației care este suficient pentru conștiență este același care susține și inteligența. Este presupunerea că inteligența și conștiența sunt interconectate intim, chiar constitutiv: conștiența va apărea, la un moment dat, pe parcurs.

Dar această presupunere nu are fundament. Așa cum am văzut în capitolul anterior, tendința de a confunda conștiența cu inteligența duce la un antropocentrism pernicios, prin care noi supra-interpretăm lumea prin prisma distorsionantă a propriilor noastre valori și experiențe. *Noi* suntem conștienți, *noi* suntem inteligenți și noi suntem atât de mândri de specia noastră și de autodeclarata noastră inteligență, încât presupunem că inteligența este legată inextricabil cu statutul nostru de conștiență și viceversa.

Deși inteligența oferă un meniu bogat de stări de conștiență ramificate pentru organismele conștiente, este o greșeală să presupunem că inteligența – cel puțin în forme avansate – este fie necesară, fie suficientă pentru conștiență.

Scanned with OKEN Scanner

Dacă insistăm pe ideea că inteligența și conștiența sunt interconectate intrinsec, ne vom grăbi să atribuim conștiență sistemelor artificiale care par să fie inteligente și să o negăm la alte sisteme – cum ar fi alte animale –, care nu satisfac standardele noastre umane chestionabile referitoare la competența cognitivă.

În ultimii ani, aceste presupuneri despre necesar și suficient au fost aruncate pe fereastră de către alte nenumărate preocupări și înțelegeri eronate, care au conferit perspectivei conștienței artificiale o urgență și o strălucire apocaliptică pe care nu le merită.

Iată câteva dintre ele. Există îngrijorarea că AI – conștien-tă sau nu – se află într-o cursă pentru dominarea inteligenței umane, reinstanțiindu-se pe sine (bootstraping itself) din-colo de înțelegerea și controlul nostru. Este așa-numita ipoteză a „singularității", popularizată de Ray Kurzweil, și bazată pe dezvoltarea fără precedent a resurselor brute de calcul din ultimele decenii. Unde ne aflăm pe această curbă exponențială? Problema cu aceste curbe – cu care ne-am obișnuit în timpul ultimei pandemii de coronavirus – este că, oriunde te-ai situa pe ele, ceea ce urmează este o pantă extrem de abruptă, iar ceea ce a trecut deja se situează pe o pantă aproape orizontală. Perspectiva locală nu oferă niciun indiciu referitor la unde te afli. Mai avem și fricile noastre prometeice, cum că propriile noastre creații se vor întoarce, într-un fel sau în altul, împotriva noastră – frici care au fost recunoscute, reîmpachetate și revândute tot nouă de mai multe filme și cărți SF. În final, ne lipsește o definiție clară a termenului de „conștiență", atunci când vine vorba de capacitățile mașinilor. Pentru unii – inclusiv cercetători AI –, orice care răspunde la o stimulare, care învață ceva sau care se comportă în așa fel încât să maximizeze răsplata sau să atingă un scop este conștient. Pentru mine, aceasta e o supra-extensie fără sens a ce anume înseamnă, în mod rezonabil, „a fi conștient".

Scanned with OKEN Scanner

Amestecați împreună toate aceste ingrediente și veți crea, astfel, premisa impresiei multora că AI conștientă e foarte aproape și că trebuie să fim foarte îngrijorați de ce o să se întâmple atunci când sosește. Posibilitatea nu trebuie ignorată complet. Dacă se dovedește că promotorii panicarzi ai singularității au dreptate, atunci chiar că trebuie să ne îngrijorăm. Dar din punctul în care ne aflăm acum, această perspectivă este extrem de improbabilă. Mult mai probabilă este o situație de tipul celei ilustrate mai jos. Aici, conștiența nu este determinată de inteligență, iar inteligența poate exista fără conștiență. Ambele vin în multe forme și ambele sunt exprimate pe multe dimensiuni diferite – însemnând că nu există doar o singură scară nici pentru conștiență, nici pentru inteligență.

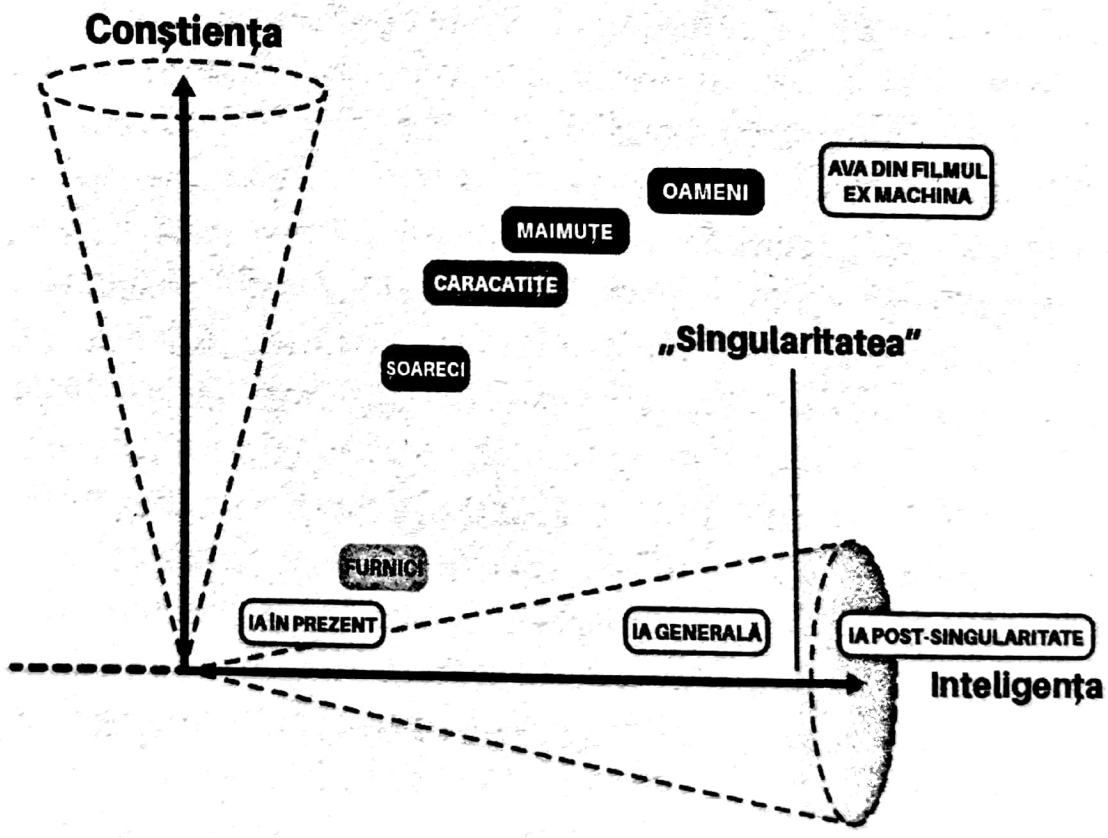

Fig. 21: Conștiența și inteligența sunt separabile și multidimensionale. Pozițiile animalelor și mașinilor (reale și imaginare) sunt ilustrative.

Scanned with OKEN Scanner

În această reprezentare, veți observa că actuala AI este localizată destul de jos pe scara inteligenței. Asta deoarece e încă neclar dacă sistemele curente de AI chiar sunt inteligente. Mare parte a AI actuale este descrisă ca fiind o recunoaștere sofisticată de către mașini – poate condimentată cu ceva planificare – a tiparelor. Inteligente sau nu, aceste sisteme fac ce fac fără a fi conștiente de nimic.

Privind în viitor, țelul ambițios al multor cercetători AI este de a dezvolta sisteme cu inteligență generală comparabilă cu a omului – așa-numita „inteligență artificială generală" sau „AI generală". Dincolo de acest punct se află *terra incognita* a inteligenței post-singularitate. Nicăieri nu există garanția că pe acest traseu se află și conștiența. Mai mult, ar putea exista mai multe forme de inteligență care să se abată de la cea umană, mai degrabă completând, decât substituind sau amplificând setul de instrumente cognitive caracteristic speciei noastre – din nou, fără o conștiență implicată.

S-ar putea ca anumite forme de inteligență să fie imposibile fără conștiență dar, chiar și așa, nu înseamnă că *toate* formele de inteligență – odată ce depășesc un prag deocamdată necunoscut – necesită conștiență. Invers, s-ar putea ca toate entitățile conștiente să fie măcar un pic inteligente, dacă inteligența este definită suficient de cuprinzător. Din nou, asta nu validează inteligența drept calea regală către conștiență.

Faptul că facem computere tot mai deștepte nu înseamnă că ele și simt. Dar asta nu înseamnă că o conștiență a mașinii ar fi imposibilă. Cum ar fi dacă am încerca să proiectăm în conștiență de la început? Dacă nu inteligența, atunci ce anume *ar fi* necesar pentru a construi o mașină conștientă?

Răspunsul la această întrebare depinde de ceea ce crezi tu că e suficient pentru un sistem ca să fie conștient, iar asta

Scanned with OKEN Scanner

depinde de acea teorie a conștienței pe care o preferi. Nu este, deci, surprinzător că există nenumărate viziuni despre ce anume ar fi nevoie pentru ca o mașină să fie conștientă.

La capătul mai relaxat al spectrului se află cei care, în acord cu funcționalismul, cred despre conștiență că este doar tipul potrivit de procesare a informației. Această procesare a informației nu trebuie să fie identică cu „inteligența", dar e, totuși, procesare a informației și, deci, este ceva ce poate fi implementat în computere. De exemplu, conform unei propuneri din revista *Science*, din 2017, se poate spune despre o mașină că este conștientă dacă procesează informația în moduri care implică „disponibilitatea globală" a informației, ceea ce ar permite „automonitorizarea" performanțelor ei. Autorii nu ne spun dacă o astfel de mașină *chiar ar fi* conștientă sau s-ar *comporta ca și cum* ar fi conștientă, dar afirmă că, pentru conștiență, nu e nevoie de nimic mai mult decât de tipul corect de procesare a informației.

O declarație mai asertivă despre mașinile conștiente a fost făcută de susținătorii teoriei informației integrate (IIT). Așa cum am văzut în capitolul 3, IIT susține despre conștiență că *este* informație integrată și că toată cantitatea de informație integrată produsă de un sistem este determinată exclusiv de către proprietățile mecanismelor sale interne – prin „structura sa cauză-efect". Conform IIT, orice mașină care generează informație integrată, indiferent din ce e făcută și indiferent de cum ar putea să arate din afară, are un anumit grad de conștiență. Cu toate acestea, IIT lasă deschisă posibilitatea ca mașinile să pară conștiente sau inteligente, sau poate ambele, din perspectiva unui observator extern, dar să aibă mecanisme care nu generează deloc informație integrată și, deci, să nu fie conștiente.

Niciuna dintre aceste teorii nu identifică inteligența cu conștiența, dar ambele susțin că mașinile care satisfac

Scanned with OKEN Scanner

anumite condiții specifice – procesare a informației de un anume tip sau informație integrată diferită de zero – ar putea fi conștiente. Dar, pentru a accepta aceste implicații, este nevoie, bineînțeles, să accepți și respectivele teorii.

Teoria animal-mașină întemeiază trăirea lumii și a sinelui pe impulsul biologic de păstrare a integrității fiziologice – de a rămâne în viață. Ce spune această teorie despre posibila existență a mașinilor conștiente?

Să ne imaginăm un robot din viitorul apropiat, cu un creier din siliciu și un corp asemănător celui uman, echipat cu tot felul de senzori și efectori. Acest robot este controlat de o rețea neuronală artificială, proiectată conform principiilor procesării predictive și a inferenței active. Semnalele care circulă prin circuitele sale implementează un model generativ al mediului său și al propriului său corp. Folosește constant acest model pentru a calcula cele mai bune predicții bayesiene despre cauzele input-urilor sale senzoriale. Aceste halucinații sintetice controlate (și care controlează) sunt direcționate, prin proiect, către menținerea robotului într-o stare de funcționare optimă – către menținerea sa, în termenii lui, „viu". Are chiar și input-uri interoceptive artificiale, semnalizând nivelurile de încărcare a bateriei, precum și integritatea motoraşelor electrice și a mușchilor săi sintetici. Cele mai bune estimări orientate-spre-control despre aceste input-uri interoceptive generează stări emoționale artificiale, care îi justifică și ghidează comportamentul.

Acest robot se comportă automat, făcând lucrurile potrivite la timpul potrivit pentru a-și îndeplini scopurile. Astfel, dă impresia, în exterior, că este un agent inteligent, simțitor. Intern, mecanismele lui se potrivesc direct pe mașinăria predictivă despre care am sugerat că susține

Scanned with OKEN Scanner

experiența umană fundamentală de întrupare și sine. Este un animal-mașină din siliciu.

Ar putea fi conștient un astfel de robot?

Răspunsul nesatisfăcător dar onest este că nu știu sigur, dar probabil nu. Teoria animal-mașină propune ideea conform căreia conștiența la om și la alte animale a apărut în decursul evoluției, se formează în fiecare dintre noi în timpul dezvoltării și operează de la un moment la altul în moduri conectate intim cu statutul nostru de sisteme vii. Toate experiențele și percepțiile noastre provin din natura noastră de mașini vii autosustenabile, preocupate de propria persistență. Părerea mea – iarăși, e doar o părere – este că materialitatea lumii se va dovedi importantă pentru toate manifestările conștienței. Unul dintre motive este acela că necesitatea reglării și autoîntreținerii în sistemele vii nu se rezumă la doar un nivel, cum ar fi integritatea întregului corp. La sistemele vii, autoîntreținerea se face și la nivel celular. Fiecare celulă din corpul nostru – din *orice* corp – își regenerează continuu condițiile care-i asigură menținerea integrității în timp. Nu putem spune același lucru pentru un computer, de acum sau din viitorul apropiat, și nici despre un animal-mașină din siliciu de tipul descris mai sus.

Nu trebuie să tragem concluzia că celulele sunt conștiente sau că toate organismele vii sunt conștiente. Ideea e că procesele de reglare fiziologică ce susțin conștiența și sinele în teoria animal-mașină sunt reinstanțiate (bootstrapped) din procesele fundamentale ale vieții care se aplică „la toate nivelurile". Din această perspectivă, nu procesarea informației, ci viul este cel care alimentează ecuațiile.

Chiar dacă mașinile cu adevărat conștiente sunt departe – asta dacă chiar ar fi posibile –, avem destule motive de

Scanned with OKEN Scanner

îngrijorare. În viitorul apropiat, este perfect plauzibil ca dezvoltarea AI și a roboticii să livreze noi tehnologii care dau impresia de conștiență, chiar dacă nu există un motiv final de a crede că *sunt* cu adevărat conștiente.

În filmul din 2014 al lui Alex Garland, *Ex Machina*, Nathan, un miliardar singuratic, geniu tech, îl invită pe Caleb, un programator excepțional, în ascunzătoarea lui îndepărtată pentru a i-o prezenta pe Ava, robotul inteligent și plin de curiozitate pe care îl crease. Sarcina lui Caleb era de a afla dacă Ava e conștientă sau e doar un robot inteligent, fără nicio lume interioară.

Ex Machina se bazează mult pe testul Turing, faimos pentru evaluarea capacității de a gândi a unei mașini. Într-o scenă incisivă, Nathan îl chestionează pe Caleb despre acest test. În versiunea standard a testului Turing, așa cum îl știe Caleb, un subiect-om interoghează atât o mașină, cât și un alt om, la distanță, doar prin mesaje scrise. O mașină trece testul atunci când subiectul-om nu reușește să distingă omul de mașină. Dar Nathan are ceva cu mult mai interesant în minte. Atunci când vine vorba de Ava, afirmă că „provocarea este să-ți *arăt* că e robot – apoi să vezi dacă mai crezi că are conștiență."

Noul joc transformă testul Turing dintr-unul de inteligență în unul de conștiență și, după cum știm deja, acestea două sunt foarte diferite. Mai mult, Garland ne arată că testul nu este cu adevărat despre robot. Așa cum explică Nathan, ceea ce contează nu e dacă Ava e mașină. Nici dacă Ava, deși e mașină, are conștiență. Ceea ce contează este dacă Ava reușește să pară, pentru o persoană conștientă, la fel de conștientă (sau nu). Extraordinar în acest schimb dintre Nathan și Caleb este faptul că dezvăluie scopul real al acestui tip de test: e un test pentru om, nu pentru mașină. Acest lucru e adevărat atât pentru testul Turing original, cât și pentru echivalentul de secol XXI orientat-spre-conștiență al lui Garland. Dialogul

Scanned with OKEN Scanner

lui Garland surprinde atât de elegant provocarea de a atribui conștiență unei mașini, încât termenul de „test al lui Garland" prinde contur – un exemplu rar de SF ce devine știință.

Despre multe programe simple de computer, inclusiv chatbots de diferite tipuri, s-a declarat că „au trecut" testul Turing, deoarece o proporție suficient de mare de subiecți a fost păcălită de suficient de multe ori. Într-un exemplu deosebit de ciudat, tot din 2014, zece din treizeci de subiecți au căzut victimă păcălelii că un chatbot care pretindea că e un băiat ucrainean, de 13 ani, era, de fapt, un băiat ucrainean real, de 13 ani. Asta a dus la proclamații zgomotoase conform cărora limitele din AI au fost, în cele din urmă, depășite. Numai că, firește, este mai ușor să joci rolul unui adolescent străin care vorbește rău engleza, decât pe al cuiva cu propria vârstă, limbă și cultură, mai ales când sunt permise doar interacțiunile la distanță prin text. Când chatbotul câștiga, răspunsul lui era următorul: „trecerea testului Turing (sic) mi se pare destul de simplă." Coborând standardele astfel, testul devine ușor de trecut. Acesta a fost un test al naivității omului, iar omul l-a picat.

Pe măsură ce AI se îmbunătățește, testul Turing ar putea fi trecut fără a fi nevoie de standarde artificiale joase. În mai 2020, laboratorul Open AI a lansat GPT-3, o vastă rețea neuronală antrenată pe exemple de limbaj natural extrase dintr-o zonă mare a internetului. Pe lângă faptul că se angajează în variate dialoguri cu chatbot, GPT-3 poate genera pasaje substanțiale de text în multe stiluri diferite, atunci când este alimentat (prompted) cu câteva cuvinte sau propoziții inițiale. Deși nu înțelege ce anume produce, fluența și stilul sofisticat al output-ului GPT-3 sunt surprinzătoare și, pentru unii, chiar înfricoșătoare. Într-un exemplu, publicat în *The Guardian*, a livrat un eseu de 500 de cuvinte despre motivele pentru care oamenii nu ar trebui să se teamă de AI – cu argumente de la psihologia

Scanned with OKEN Scanner

violenței umane până la revoluția industrială și incluzând fraza deconcertantă „AI nu ar trebui să piardă vremea încercând să înțeleagă punctele de vedere ale celor care nu au încredere în inteligența artificială".

Oricât de sofisticat ar fi, sunt destul de sigur că GPT-3 poate fi surprins de orice interlocutor om rezonabil de sofisticat. Asta s-ar putea să nu mai fie adevărat pentru GPT-4 sau GPT-10. Chiar dacă un sistem viitor, asemănător GPT, trece repetat testul Turing, va afișa mai degrabă o formă de inteligență (simulată) foarte îngustă – schimb lingvistic neîntrupat/disembodied linguistic exchange – decât inteligența naturală întrupată, „care face ceea ce trebuie atunci când trebuie", și pe care o vedem la oameni și la multe alte animale – precum și la al meu animal-mașină ipotetic din siliciu.

Atunci când ne referim la conștiență, nu există un echivalent al chatbot-ului ucrainean, darămite al GPT-oricare. Testul Garland rămâne impecabil. De fapt, tentativele de a crea simulacre de ființe simțitoare au generat, mai degrabă și adeseori, sentimente de anxietate și repulsie, decât amestecul de atracție, empatie și milă pe care Caleb îl simțea pentru Ava din *Ex Machina*.

Specialistul japonez în robotică, Hiroshi Ishiguro, a petrecut zeci de ani construind roboți care să semene cât de mult se poate cu ființele umane. Îi numește „Geminoizi". Ishiguro a creat copii Geminoid ale lui (pagina alăturată) și ale fiicei lui (care avea 6 ani la acea vreme), precum și o prezentatoare Geminoid nipono-europeană pe baza unui amestec de circa 30 de persoane diferite. Fiecare Geminoid este construit prin scanarea 3D detaliată a corpului și are motorașe pneumatice capabile să genereze o mare varietate de expresii faciale și gesturi. Fiecare dintre aceste

Scanned with OKEN Scanner

dispozitive este un AI miniatural – sunt, toate, destinate imitării omului, cu posibile aplicații, printre altele, în prezența la distanță sau „teleprezență". Ishiguro și-a folosit, la un moment dat, Geminoidul pentru a ține o prelegere la distanță de 45 de minute în fața a 150 de studenți.

Geminoizii sunt, fără îndoială, înfiorători. Sunt foarte aproape de realitate, dar nu suficient. Gândiți-vă la întâlnirea cu un Geminoid ca la inversul întâlnirii cu o pisică. Cu o pisică – sau chiar și o caracatiță – există o senzație imediată de prezență a unei alte entități conștiente, chiar dacă felul în care arată este diferit. Cu un Geminoid, asemănarea fizică izbitoare, dar imperfectă, accentuează senzațiile de alteritate și de lipsă a conectării. Un studiu din 2009 a arătat că ceea ce simțeau cei care întâlneau un Geminoid era predominant frică.

Fig. 22: Hiroshi Ishiguro cu Geminoidul său.
Cu permisiunea ATR Hiroshi Ishiguro Laboratories.

Scanned with OKEN Scanner

Acest tip de reacție exemplifică așa-numita „vale stranie“ (uncanny valley), un concept formulat de un un alt cercetător japonez, Masahiro Mori, în 1970. Mori a sugerat că, pe măsură ce un robot începe să arate ca un om, va determina reacții empatice și tot mai pozitive din partea oamenilor (ca C-3PO din *Star Wars*). Odată ce trece de un anume punct, însă, în care ajunge să semene foarte mult cu un om în anumite privințe, dar nu și în altele, aceste reacții coboară rapid în zona repulsiei și fricii – valea stranie –, revenind la loc pe măsură ce asemănarea ajunge până la confuzie. S-au formulat multe teorii pentru a explica existența văii stranii, ceea ce nu se poate nega este însăși existența ei.

Deși roboții din lumea reală scapă cu greu din valea stranie, evoluțiile din lumea virtuală le-au permis să fie deja pe panta ascendentă, ba chiar să fi trecut de ea. Evoluția recentă din machine learning pe baza „Rețele neuronale adversariale generative“ – GANN = generative adversial neural network – poate genera fețe fotorealiste de oameni care nu au existat vreodată (vezi mai jos).* Aceste imagini sunt create prin amestecarea trăsăturilor dintr-o mare bază de date de fețe reale, implicând tehnici similare celor

Fig. 23: Opt fețe. Aceste persoane nu sunt adevărate.

* Aceste fețe artificiale au fost generate folosind thispersondoesnotexist.com.

Scanned with OKEN Scanner

pe care le folosim noi în mașina de halucinat (descrisă în capitolul 6). Combinate cu tehnologii „deepfake", care pot anima aceste fețe, făcându-le să și vorbească, iar când ceea ce spun este amplificat de un software tot mai sofisticat de recunoaștere a discursului și de producere a limbajului, cum este GPT-3, generează brusc o lume plină de reprezentări virtuale ale oamenilor reali. În această lume, ne vom obișnui cu faptul că nu vom putea să știm cine e real și cine nu.

Oricine crede că aceste evoluții vor atinge un plafon înainte ca un test Turing îmbunătățit video să fie trecut convingător, se înșală. Un astfel de mod de a gândi dezvăluie fie un caz rezistent de excepționalism uman, fie un eșec al imaginației, fie ambele. Se va întâmpla. Rămân două întrebări. Prima este dacă aceste creații virtuale noi vor fi capabile să treacă în lumea reală, traversând valea stranie în care Geminoizii lui Ishiguro au rămas captivi. A doua este dacă testul Garland va eșua și el. Vom simți că aceste noi creaturi sunt, de fapt, și conștiente și inteligente – chiar și când știm că nu sunt nimic mai mult decât linii de cod?

Și dacă simțim astfel, cum anume ne va influența?

Avansul rapid al AI – indiferent de amestecul de exagerare și de realitate care o alimentează – a stârnit o discuție tot mai necesară despre etică. Multe îngrijorări de ordin etic au de a face cu consecințele economice și societale ale tehnologiilor din viitorul apropiat, precum mașinile care se conduc singure și muncitorii automați din fabrici, unde perturbările semnificative sunt inevitabile.* Există îngrijorări legitime referitoare la delegarea capacității de

* Unele dinte aceste tehnologii nu sunt chiar atât de noi cum par. Colegul meu, Ryota Kanai, a formulat recent fraza „un cal este, practic, un cal care se autoconduce".

Scanned with OKEN Scanner

decizie unui sistem artificial, ale cărui acțiuni interioare pot fi susceptibile la tot felul de prejudecăți și capricii și care pot rămâne opace – nu numai pentru cei afectați, dar și pentru cei care l-au proiectat. La capătul extrem al spectrului, ce orori s-ar putea dezlănțui dacă un sistem AI ar ajunge responsabil de armele nucleare sau de coloana vertebrală a internetului?

Există și îngrijorări etice referitoare la consecințele psihologice și comportamentale ale AI și machine learning. Invazia vieții private prin deepfakes, modificările comportamentului prin algoritmi predictivi, distorsiunea convingerilor în bule filtrate și „camere ecou" (echo chambers) din social media sunt doar câteva dintre nenumăratele forțe care tensionează structura societăților noastre. Prin eliberarea acestor forțe, ne cedăm de bunăvoie identitățile și autonomia unor corporații de date fără chip într-un vast experiment global necontrolat.

Pe acest fond, discuțiile etice despre conștiența mașinii pot părea indulgente și obscure. Numai că nu sunt așa. Discuțiile acestea devin necesare, chiar dacă respectivele mașini nu au (deocamdată) conștiență. Atunci când se va trece testul Garland, vom trăi împreună cu entități despre care *simțim* că au propriile vieți interioare, chiar dacă am putea ști sau crede că nu au. Consecințele psihologice și comportamentale sunt greu de estimat. O posibilitate este aceea că vom învăța să diferențiem cum *simțim* de cum ar trebui să *acționăm*, astfel încât va părea natural să ne pese de un om, dar nu și de un robot, chiar dacă simțim că ambii au conștiență. Nu e clar cum anume ne va influența așa ceva psihologia individuală.

În seria TV *Westworld*, roboți foarte asemănători oamenilor sunt proiectați special pentru a fi abuzați, omorâți, violați – servind ca defulare pentru comportamentele cele mai depravate ale omenirii. Ar putea fi posibil să torturezi un robot despre care simți că e conștient, dar știi, în

Scanned with OKEN Scanner

același timp, că nu e, fără ca mintea să ți se fractureze? Cu mințile de acum, astfel de comportamente ar fi în zona înalt-sociopată. O altă posibilitate este aceea ca cercul preocupării noastre morale să fie distorsionat de tendința noastră antropocentrică de a simți empatie mai mare pentru entități cu care ne simțim mai asemănători. În acest scenariu, s-ar putea să ne pese mai mult de gemenii Geminoid din următoarea generație decât de ceilalți oameni, nemaivorbind de alte animale.

Bineînțeles, nu toate scenariile viitorului trebuie să fie distopice. Dar, pe măsură ce întrecerea dintre progres și moda AI capătă ritm, etica informată, aliată psihologic, trebuie să-și joace rolul. Nu e bine, pur și simplu, să pui la dispoziție tehnologie nouă și să aștepți să vezi ce se întâmplă. Mai presus de toate, obiectivul standard al AI de a recrea și depăși inteligența umană nu ar trebui urmărit orbește. Așa cum a spus, cu înțelepciune, Daniel Dennet, „construim unelte inteligente, nu colegi" și trebuie să fim siguri că putem recunoaște diferența.

Apoi vine posibilitatea unei adevărate conștiențe a mașinii. Dacă am introduce – deliberat sau nu – noi forme de experiență subiectivă, ne vom confrunta cu o criză etică și morală de proporții fără precedent. Orice are statut de conștient are și statut moral. Vom fi obligați să-i minimizăm potențialul de suferință în același fel în care suntem obligați să minimizăm suferința ființelor vii – și nu suntem prea buni la asta. Iar pentru acești agenți (agents) artificiali, posibil simțitori, apare o provocare suplimentară – s-ar putea să n-avem nicio idee despre tipurile lor de conștiență. Să ne imaginăm un sistem supus unei noi forme de suferință, pentru care noi, oamenii, nu avem echivalent sau concept, nici instincte prin care să o recunoaștem. Să ne imaginăm un sistem pentru care diferența dintre emoțiile pozitive și negative nici nu se aplică, pentru care nu există dimensiune fenomenologică

Scanned with OKEN Scanner

corespunzătoare. Aici, provocarea etică constă în aceea că nici nu vom ști care sunt problemele etice relevante.

Oricât de departe ar rămâne conștiența artificială, chiar și posibilitatea aceasta îndepărtată ar trebui luată în considerare. Deși nu știm ce ar fi nevoie pentru a crea o mașină conștientă, nu știm nici ce *nu* ar fi nevoie.

În iunie 2019, filosoful german Thomas Metzinger a solicitat un moratoriu imediat, de 30 de ani, asupra întregii cercetări destinate generării a ceea ce el a numit „fenomenologie sintetică", tocmai din aceste motive. Eram acolo atunci când a făcut acest anunț. Fusesem amândoi invitați să luăm cuvântul la întâlnirea pentru conștiența artificială, găzduită de Centrul Leverhulme pentru viitorul inteligenței, din Cambridge. Solicitarea lui Metzinger e greu de urmat la literă, deoarece mare parte, dacă nu toată modelarea computerizată din psihologie ar putea intra în categoria menționată de el, dar sensul mesajului e clar. Nu ar trebui să continuăm în a crea conștiență artificială doar pentru că credem că e interesant, util sau cool. Cea mai bună etică este cea preventivă.

În perioada de glorie a vitalismului, ar fi părut absurd să vorbim de etica vieții artificiale așa cum ni se poate părea azi despre etica conștienței artificiale. Dar iată-ne, cu puțin peste 100 de ani mai târziu, nu numai cu o înțelegere profundă a ceea ce face viața posibilă, dar și cu nenumărate ustensile noi cu care o modificăm și chiar o creăm. Avem tehnici de editare genetică precum CRISPR, care permite cercetătorilor să modifice cu ușurință secvențele de ADN și să schimbe funcția genelor. Avem chiar și capabilitatea de a dezvolta organisme în întregime sintetice, generate din „gene up": în 2019, cercetătorii de la Cambridge au creat o variantă a *Escherichia coli*, cu un genom complet sintetic. Etica creării de noi forme de viață devine brusc extrem de relevantă.

Scanned with OKEN Scanner

Și poate că mai degrabă biotehnologia, nu neapărat AI, ne va duce mai aproape de conștiența sintetică. Aici, „organoidele cerebrale" capătă semnificație specială. Acestea sunt structuri extrem de mici, asemănătoare creierului, alcătuite din neuroni reali, crescute din celule stem pluripotente (celule care se pot diferenția în nenumărate forme). Deși nu sunt „mini creieri", organoidele cerebrale seamănă cu creierul uman în dezvoltare în feluri care le fac utile ca modele de laborator pentru afecțiunile medicale în care dezvoltarea creierului se face eronat. Pot aceste organoide să găzduiască o formă primitivă de conștiență fără corp? E greu de exclus această posibilitate, mai ales atunci când încep să manifeste unde coordonate de activitate electrică, nu foarte diferite de cele observate la copiii prematuri, după cum a constatat un studiu recent.

Spre deosebire de computere, organoidele cerebrale sunt făcute din același material fizic ca și creierii adevărați, înlăturând, astfel, unul dintre obstacolele care ne-ar împiedica să credem despre ele că ar putea fi conștiente. Pe de altă parte, rămân extrem de simple, sunt complet în afara unui corp și nu interacționează deloc cu lumea exterioară (deși e posibil să le cablăm la camere și brațe robotice și alte asemenea). În ce mă privește, în timp ce organoidele curente sunt departe de a fi conștiente, întrebarea va rămâne deconcertant de deschisă, pe măsură ce tehnologia se dezvoltă. Ne reamintesc de nevoia de etică preventivă. Posibilitatea existenței unei conștiențe a organoidului urgentează etica nu numai pentru că nu poate fi exclusă, ci și din cauza amplorii potențial implicate. Așa cum spunea cercetătoarea Alysson Muotri, „vrem să facem ferme de astfel de organoide".

Scanned with OKEN Scanner

De ce este așa de ispititoare perspectiva conștienței mașinii? De ce exercită o astfel de atracție asupra imaginației noastre colective? Am ajuns să cred că are de a face cu un fel de tehno-răpire, o dorință profundă de a transcende existența noastră biologică, materială, circumscrisă și dezordonată, pe măsură ce se apropie sfârșitul. Dacă mașinile conștiente ar fi posibile, cu ele vine și posibilitatea de a ne reloca mințile conștiente din sediul lor organic (wet-ware) în circuitele impecabile ale unui supercomputer viitor, care nu îmbătrânește și nu moare niciodată. Acesta este teritoriul mind uploading/ încărcarea minții (pe un cip, n. trad.), un refren preferat al futuriștilor și transhumaniștilor, pentru care o viață nu e suficientă.

Unii cred că suntem deja acolo. Filosoful Nick Bostrom, de la Universitatea Oxford, autorul „argumentului simulării", prezintă un caz statistic ce propune ideea că avem mai multe șanse să fim parte a unei simulări computerizate extrem de sofisticate, proiectate și implementate de descendenții noștri, superiori tehnologic și obsedați genealogic, decât să fim parte a rasei umane biologice, originale. Din această perspectivă, suntem deja agenți simțitori virtuali într-un univers virtual.

Cei pasionați de tehno-răpire văd cum se apropie rapid o singularitate, un punct critic al istoriei în care AI este gata să se reinstanțieze (bootstrap) pe sine dincolo de înțelegerea și de controlul nostru. Într-o lume post-singularitate, mașinile conștiente și simulările strămoșilor abundă. Noi, forme de viață pe bază de carbon, vom rămâne mult în urmă, momentul nostru sub soare s-a sfârșit.

Nu e nevoie de prea multă introspecție sociologică pentru a constata preferința elitei noastre tehnologice pentru această licoare amețitoare. În astfel de circumstanțe, devine pivot al unei tranziții fără precedent în istoria umanității, primind drept premiu nemurirea. Asta

Scanned with OKEN Scanner

se întâmplă atunci când excepționalismul uman sare complet de pe șine. Văzut astfel, tot zgomotul despre conștiența mașinii este simptomatic pentru o alienare în creștere față de natura noastră biologică și de zestrea noastră evolutivă.

Perspectiva animal-mașină diferă de acest narativ în aproape toate felurile. În teoria mea, așa cum am văzut, întregul experienței umane și al vieții mentale rezultă ca urmare a, nu împotriva naturii organismelor biologice care se autosusțin și cărora le pasă de propria persistență. Această perspectivă asupra conștienței și a naturii umane nu exclude posibilitatea existenței mașinii conștiente, dar subminează narațiunea tehno-răpirii computerelor-curând-simțitoare, care ne alimentează temerile și ne bântuie visele. Din perspectiva animalului-mașină, nevoia de a înțelege conștiența ne plasează tot mai mult în natură, nu departe de ea.

Exact așa cum ar trebui.

Nota trad.:
chatbot = program care mimează conversația umană prin scris sau voce;
deepfake = hipertrucaj.

Scanned with OKEN Scanner

Epilog

Vreau să am control,
Vreau un corp perfect,
Vreau un suflet perfect.

RADIOHEAD, „Creep" (1992)

În ianuarie 2019, am ajuns, pentru prima dată, față în față cu un creier uman viu. Douăzeci de ani fascinanți după ce am început cercetarea în știința conștienței, zece ani de când laboratorul nostru din Sussex și-a deschis porțile, trei ani după propria mea anestezie – uitarea indusă de la începutul cărții. După tot acest timp, privind la suprafața gri-albă, pulsând ușor, a cortexului, străbătută de vase delicate cu sânge de un roșu închis, părea, din nou, de neconceput ca o grămăjoară de așa ceva să dea naștere unui întreg univers de gânduri, simțiri, percepții – unei vieți trăite din plin, la persoana I. Sentimentul profund de uimire pe care l-am trăit s-a amestecat inconfortabil în mintea mea cu vechea glumă care spune că un transplant de creier e singura operație în care e bine să fii donatorul, nu primitorul.

Eram la Michael Carter, specialist în neurochirurgie pediatrică la spitalul pentru copii Bristol Royal din vestul Angliei. Mă invitase să observ unele dintre cele mai serioase proceduri de neurochirurgie realizate vreodată oriunde. Pacientul, un copil de doar 6 ani, fusese programat pentru emisferotomie. Suferea de epilepsie severă încă de la naștere. Crizele începeau în emisfera cerebrală dreaptă, care fusese foarte deteriorată în cursul nașterii lui premature. Toate medicamentele standard antiepileptice eșuaseră, astfel că s-a apelat la un neurochirurg.

Scanned with OKEN Scanner

Emisferotomia implică, în acest caz, deconectarea neuronală totală de emisfera dreaptă, disfuncțională. Chirurgul pătrunde în creier prin partea dreaptă, înlătură („rezecă") lobul temporal, apoi face o tăietură prin toate fasciculele de conexiuni – tracturile de materie albă – care leagă emisfera dreaptă de restul creierului și al corpului. Emisfera izolată rămâne în interiorul craniului, încă conectată la rețeaua vasculară. Este o insulă cerebrală, vie, dar izolată. O versiune extremă a operației mai familiare, de separare a emisferelor cerebrale, deconectarea neuronală completă previne răspândirea în întreg creierul a furtunilor electrice care pornesc în emisfera dreaptă, cea afectată. Dacă operația e făcută suficient de devreme, creierul tânăr este suficient de adaptabil pentru ca emisfera rămasă funcțională să preia mare parte, dacă nu toate sarcinile. În ciuda naturii radicale a acestei proceduri chirurgicale, deși fiecare caz e diferit, rezultatele sunt, în general, bune.

Această procedură chirurgicală a început pe la prânz și a durat puțin peste opt ore. Eu nu pot lucra mai mult de 5 minute fără să mă distragă ceva – un email, scorul la cricket, o altă cană de ceai. Michael a lucrat fără pauză oră după oră, răbdător, metodic, fără oprire, susținut de un stagiar neurochirurg și o echipă de asistente care se tot schimbau. Pe la jumătatea timpului, când stagiarul a luat o scurtă pauză, am fost invitat în sala de operație, să privesc prin microscopul chirurgical. Nu mă așteptam la un asemenea privilegiu. Privind în cavitățile cerebrale puternic luminate ale copilului, am încercat să-mi verific cunoașterea abstractă despre diferitele regiuni și căi din țesuturile expuse în fața mea. Nu avea aproape niciun sens. Ierarhiile corticale precise și valurile de semnale care curg bottom-up și top-down nu erau la vedere. Creierul devenise, brusc, de nepătruns și rămăsesem uluit și de profesionalismul neurochirurgului și de realitatea materială a acestui organ magic. Era aproape transcendental. Se trăsese o cortină,

Scanned with OKEN Scanner

dezvăluind ceva prea intim pentru a fi expus în mod obișnuit. Priveam direct în mecanica sinelui uman.

Procedura chirurgicală a decurs conform planului. La ceva timp după ora 8, Michael l-a lăsat pe stagiar să termine de cusut scalpul și m-a luat cu el să îmi prezinte familia copilului. Erau recunoscători și ușurați. Mă întrebam ce ar fi simțit ei dacă ar fi văzut ce am văzut eu în acea zi.

Mai târziu, conducând spre casă prin întunericul iernii, gândurile mi s-au întors la descrierea pe care David Chalmers a făcut-o despre problema grea a conștienței: „există un consens general în ceea ce privește originea experienței dintr-o bază fizică, dar nu avem o explicație suficient de bună pentru de ce și cum se formează experiența. De ce ar genera procesarea fizică o viață interioară atât de bogată, până la urmă? Faptul că se întâmplă pare nerezonabil obiectiv, dar totuși se întâmplă".

Confruntată cu acest mister, filosofia a oferit mai multe variante, de la panpsihism (conștiență mai mult sau mai puțin ubicuă) la materialism eliminativ (fără conștiență, cel puțin nu așa cum o știm) și orice între cele două. Dar știința conștienței nu este despre a alege dintr-un meniu, indiferent de cât de sofisticat e restaurantul sau cât de talentat e maestrul bucătar. Este, mai degrabă, mai aproape de gătitul cu ce ai în frigider, situație în care combini și recombini bucățele și ingrediente din filosofie, neuroștiință, psihologie, informatică (computer science), psihiatrie, învățare automată (machine learning) etc., pentru a crea ceva nou.

Aceasta este esența abordării conștienței prin prisma problemei reale. Accepți că există, apoi te întrebi cum anume sunt structurate diferitele proprietăți fenomenologice ale conștienței – care s-ar traduce prin cum anume

Scanned with OKEN Scanner

sunt structurate experiențele conștiente, ce forme iau – relativ la proprietățile creierului, creier care e încastrat în corp și în diverse lumi. Răspunsurile la aceste întrebări pot fi formulate prin identificarea corelațiilor dintre un tip sau altul de experiență conștientă, dar nu trebuie să se oprească aici. Provocarea e să construim punți explicative tot mai solide între mecanism și fenomenologie, astfel ca relațiile pe care le identificăm să nu fie arbitrare, ci să aibă un înțeles. Ce înseamnă „a avea înțeles", în acest context? Din nou: explicație, predicție și control.

Din punct de vedere istoric, această strategie reflectă modul în care înțelegerea noastră științifică asupra lumii a transcens gândirea magică a vitalismului, prin individualizarea proprietăților sistemelor vii și, apoi, prin explicarea fiecăreia în termeni de mecanisme care le susțin. Viul și conștiența sunt, firește, lucruri diferite, deși sper că am convins, până acum, că sunt mult mai intim conectate decât par la prima vedere. Oricare ar fi calea, strategia este aceeași. Dacă în loc să punem în frunte rezolvarea problemei grele a conștienței, evitând să luăm în considerare toate calitățile experiențiale ale conștienței, am aborda problema reală, atunci ar exista o speranță autenică de reconciliere între fizic și fenomenal – dizolvând, nu rezolvând problema.

Am început cu *nivelul* de conștiență – diferența dintre a fi în comă și a fi complet treaz –, unde ne-am concentrat pe importanța *măsurătorilor*. Ideea esențială aici este că măsurătorile alese, precum densitatea cauzală și informația integrată, nu sunt arbitrare. Mai degrabă, acestea surprind proprietăți foarte bine conservate ale *tuturor* experiențelor conștiente, și anume că fiecare experiență conștientă este simultan unificată cu, dar și distinctă de toate celelalte experiențe conștiente. Fiecare episod conștient este trăit „dintr-o bucată", iar fiecare experiență este așa cum este, nu altfel.

Scanned with OKEN Scanner

Am trecut apoi la natura *conținutului* conștient, în special a experienței de a fi un *sine* conștient. Am venit cu mai multe provocări la modul în care par lucrurile astfel că, în fiecare caz, am stimulat adoptarea de perspective noi, post-copernicane, asupra percepției conștiente.

Prima provocare a fost cea de a înțelege percepția mai degrabă ca pe o construcție activă, orientată spre acțiune, decât ca pe o înregistrare pasivă a unei realități externe, obiective. Lumile pe care le percepem sunt mai puțin, dar și mai mult decât ceea ce ar putea fi această realitate externă obiectivă. Creierul nostru creează lumile noastre prin procese de cea mai bună estimare bayesiană, în care semnalele senzoriale servesc, în primul rând, la controlul ipotezelor noastre perceptive în continuă schimbare. Trăim într-o halucinație controlată pe care evoluția a proiectat-o pentru utilitate, nu pentru acuratețe.

A doua provocare a întors invers această perspectivă, către experiența de a fi un sine. Am explorat cum anume sinele este el însuși o percepție, un alt tip de halucinație controlată. De la experiențe ale identității personale și ale continuității în timp, înapoi către simțul rudimentar de a fi pur și simplu un corp viu, toate aceste piese ale sinelui depind de același dans delicat între predicțiile perceptuale inside-out/dintre interior spre afară și erorile de predicție outside-in/dinspre exterior spre interior, chiar dacă cea mai mare parte a acestui dans are loc între limitele corpului.

Provocarea finală a fost de a vedea dacă mașinăria predictivă a percepției conștiente își are originea și funcția primară nu în reprezentarea lumii corpului, ci în controlul și reglarea stărilor noastre fiziologice. Suma tuturor percepțiilor și cognițiilor noastre – întreaga panoramă a experienței umane și a vieții mentale – este sculptată de impulsul biologic profund de a rămâne în viață. Percepem lumea din jurul nostru și pe noi în ea *cu*, *prin* și *datorită* corpurilor noastre vii.

Scanned with OKEN Scanner

Aceasta este teoria mea despre animalul-mașină, versiunea de secol XXI – sau inversiunea – conceptului de *homme machine*/om mașină a lui Julien Offray de La Mettrie. Aici se produc schimbările cele mai semnificative în modul de gândire despre conștiență și sine.

Există această idee conform căreia experiențele de „a fi un sine" sunt foarte diferite de trăirea lumii din jurul nostru. Acum le putem înțelege ca pe diferite expresii ale acelorași principii de predicție perceptivă, cu diferențele de fenomenologie generate de diferențele dintre tipurile de predicție implicate. Unele inferențe perceptive sunt direcționate spre găsirea de obiecte în afara noastră, în timp ce altele se referă în exclusivitate la controlul interiorului corpului.

Legând viețile noastre mentale de realitatea noastră fiziologică, conceptele vechi de când lumea despre continuitatea dintre viu și minte capătă substanță, susținută de coloane solide ale procesării predictive și ale principiului energiei libere. Iar continuitatea profundă ne permite, în schimb, să ne vedem în relație mai apropiată cu alte animale și cu restul naturii, dar și la depărtare echivalentă de organizarea imaterială a AI. Pe măsură ce conștiența și viul se adună împreună, conștiența și inteligența se îndepărtează una de alta. Această reorientare a locului nostru în natură se aplică nu numai corpurilor noastre fizice, biologice, ci și minților noastre conștiente, felului în care trăim lumea din jurul nostru și pe noi înșine.

Ori de câte ori știința ne-a scos din centrul a orice, ne-a răsplătit cu mult mai mult, în cele din urmă. Revoluția copernicană ne-a dăruit un univers – pe care descoperirile astronomice din ultima sută de ani l-au extins mult peste limitele imaginației umane. Teoria evoluției prin selecție

Scanned with OKEN Scanner

naturală, a lui Charles Darwin, ne-a dăruit o familie, o conexiune cu toate celelalte specii de viețuitoare și o apreciere a timpului și a forței designului evolutiv. Iar acum știința conștienței, din care teoria animal-mașină este doar o parte, sparge ultimul bastion al excepționalismului uman – statutul presupus „special" al minților noastre conștiente – arătându-ne că și acesta este adânc întipărit în mulțimea de tipare ale naturii.

Absolut orice din experiența conștientă este o percepție de un fel, fiecare percepție este un fel de halucinație controlată – sau care ne controlează. Ce mă încântă cel mai mult la acest mod de gândire este cât de departe ne poate duce. Experiența liberului arbitru este percepție. Curgerea timpului este percepție. Poate și structura tridimensională a lumii pe care o trăim și impresia despre conținutul experienței perceptive că e real obiectivă – acestea ar putea fi tot aspecte ale percepției. Uneltele științei conștienței ne permit să ne apropiem de noumen-ul lui Kant, realitatea supremă, de neștiut, din care facem și noi parte. Toate aceste idei sunt testabile și, indiferent de datele obținute, doar punând întrebări de acest fel ne remodelează înțelegerea despre ce anume este conștiența, cum are loc și la ce folosește. Fiecare etapă îndepărtează impresia ispititoare, dar inutilă, despre conștiență că ar fi un singur lucru – un mare mister înfricoșător în căutarea unei soluții pe măsură.

Există și nenumărate implicații practice. Măsurători ale nivelului de conștiență inspirate din teorie inaugurează conștiento-„metre" tot mai capabile să detecteze conștiență reziduală – „ascunsă" – la pacienții care nu mai răspund comportamental. Modelele computaționale ale percepției predictive aruncă o nouă lumină asupra bazelor halucinației și delirului, inaugurând o transformare a psihiatriei – trecerea de la tratarea simptomelor la descoperirea cauzelor. Și există tot felul de noi direcții pentru AI, interfețe creier-mașină și realitate virtuală,

Scanned with OKEN Scanner

printre alte nenumărate tehnologii, noi sau deja implementate. Căutarea bazei biologice a conștienței este surprinzător de utilă.

Acestea fiind spuse, contemplarea misterului conștienței este – și întotdeauna va fi – o călătorie personală profundă. La ce ar fi bună o știință a conștienței dacă nu ar arunca o nouă lumină asupra vieților mentale individuale și asupra vieților interioare ale celor din jurul nostru?

Aceasta este adevărata promisiune a adevăratei probleme. Indiferent de unde ne duce, acest drum ne va ajuta să înțelegem extrem de multe despre trăirea conștientă a lumii din jurul nostru și a noastră, în interiorul ei. Vom vedea cum universul nostru interior este parte a, nu separat de restul naturii. Și, deși poate nu ne gândim la asta atât de des pe cât ar trebui, vom avea șansa de a face pace cu tot ceea ce se întâmplă – sau nu se întâmplă – atunci când halucinația controlată de *a fi tu* se destramă în nimic. Atunci când uitarea nu este o întrerupere indusă de anestezie a fluxului conștienței, ci o întoarcere în eternitatea din care fiecare dintre noi am apărut la un moment dat.

La finalul *acestei* istorisiri, când viața la persoana I ajunge la propriile concluzii, poate că nu e chiar așa de rău să rămână și un pic de mister.

Scanned with OKEN Scanner

Mulțumiri

Ideile din această carte au fost modelate prin nenumărate discuții cu prieteni, colegi, studenți, profesori și mentori timp de mai mult de două decenii. Le mulțumesc tuturor.

Mulțumesc tuturor membrilor – foști și actuali – grupului mei de cercetare de la Sussex. A fost și continuă să fie un privilegiu să lucrez cu voi. Sunt recunoscător în special celor din grup a căror cercetare este descrisă în această carte. Mulțumesc, Adam Barrett, Lionel Barnett, Peter Lush, Alberto Mariola, Yair Pinto, Warrick Roseboom, Michael Schartner, David Schwartzman, Maxine Sherman, Keisuke Suzuki și Alexander Tschantz.

Mulțumesc, de asemenea, Manuel Baltieri, Reny Baykova, Luc Berthouze, Daniel Bor, Chris Buckley, Acer Chang, Paul Chorley, Ron Chrisley, Andy Clark, Marianne Cole, Clémence Compain, Guillaume Corlouer, Hugo Critchley, Zoltán Dienes, Tom Froese, Paul Graham, Inman Harvey, Owen Holland, Ryota Kanai, Tomasz Korbak, Isabel Maranhão, Federico Micheli, Beren Millidge, Thomas Nowotny, Andy Philippides, Charlotte Rae, Colin Reveley, Ryan Scott, Lina Skora, Nadine Spychala, Marta Suarez-Pinilla, Chris Thornton, Hao-Ting Wang și Jamie Ward – colegi și prieteni din trecutul și prezentul Sussex, ale căror muncă și idei au contribuit atât de mult la ale mele, și dintre care mulți au fost atât de buni să citească și să comenteze diverse părți ale cărții.

Sunt extrem de recunoscător că am avut ca mentor un om de știință remarcabil în fiecare etapă a carierei mele. Nicholas Mackintosh, care s-a stins în 2015, m-a remarcat în timpul studenției la Cambridge și mi-a dat încrederea

Scanned with OKEN Scanner

să mă stabilesc în lumea academică. Phil Husbands mi-a supervizat DPhil (ceea ce se numește PhD la Sussex), acordându-mi libertatea de a explora în timp ce veghea să rămân pe drumul cel bun. Gerald Edelman, care s-a stins în 2014, mi-a fost mentor mai mult de 6 ani, pe vremea perioadei postdoctorale. Sub îndrumarea lui, interesul pentru conștiență a devenit subiectul cercetării mele. Aș vrea să aduc mulțumiri și pentru Margaret Boden, Andy Clark, Daniel Dennett și Thomas Metzinger, care au fost influențe extraordinare și inspirații în decursul multor ani.

Mai multe persoane și-au asumat sarcina de a citi și comenta în detaliu primele propuneri complete ale cărții. Mulțumiri din inimă se cuviņ, deci, pentru Tim Bayne, Andy Clark (din nou), Claudia Fischer, Jakob Hohwy și Murray Shanahan. Ideile și sugestiile lor m-au ajutat enorm. Sunt, de asemenea, extrem de recunoscător – Karl Friston, Marcello Massimini, Thomas Metzinger (din nou), Aniruddh (Ani) Patel și Adrian Owen – pentru sfaturile detaliate pe care mi le-ați dat în zonele selectate. Multe mulțumiri, Steve West (Lazy Chief) pentru ilustrațiile stilate. Mulțumesc, Baba Brinkman, al cărui cadou despre cuvinte este o permanentă inspirație. Lucrul cu el la al său *Rap Guide to Consciousness* a fost o adevărată iluminare. Mulțumesc, Michael Carter, pentru că m-ai adus în sala de operație, pentru a fi martor la o emisferotomie. Mulțumesc, David Edelman și Graziano Fiorito, pentru că m-ați adus în lumea caracatițelor, Ian Cheney și Laurie Santos, pentru că mi-ați arătat lumea maimuțelor, Luise Schreiter, pentru că m-ai adus în altă lume.

Dintre mulți alții ale căror sfaturi, idei și susținere mi-au fost de ajutor în toți acești ani, voi menționa doar câțiva. Mulțumesc, Anil Ananthaswamy, Chris Anderson, Bernard Baars, Lisa Feldman Barrett, Isabel Behncke,

Scanned with OKEN Scanner

Tristan Bekinschtein, Yoshua Bengio, Heather Berlin, Matt Bergman, David Biello, Robin Carhart-Harris, Olivia Carter, David Chalmers, Craig Chapman, Axel Cleeremans, Athena Demertzi, Steve Fleming, Zaferios Fountas, Friday Football, Chris Frith, Uta Frith, Alex Garland, Mariana Garza, Mel Goodale, Annaka Harris, Sam Harris, Nick Humphrey, Rob Illife, Robin Ince, John Iversen, Eugene Izhikevich, Alexis Johansen, Robert Kentridge, Christof Koch, Sid Kouider, Jeff Krichmar, Victor Lamme, Hakwan Lau, Steven Laureys, Rafael Malach, Daanish Masood, Simon McGregor, Pedro Mediano, Lucia Melloni, Liad Mudrick, Phil Newman, Angus Nisbet, the Parasites, Megan Peters, Giovanni Pezzulo, Tony Prescott, Blake Richards, Fernando Rosas, Adam Rutherford, Tim Satterthwaite, Tom Smith, Narayanan Srinivasan, Catherine Tallon-Baudry, Giulio Tononi, Nao Tsuchiya, Nick Turk-Browne, Lucina Uddin, Simon van Gaal, Bruno van Swinderen, Anniek Verholt, Paul Verschure, Lucy Walker, Nigel Warburton, Lisa Westbury și Martijn Wokke.

Acești oameni m-au ajutat să țin drumul drept, în toate modurile posibile, de-a lungul anilor. Asta nu înseamnă că au fost de acord cu ceea ce afirmam. Orice eroare sau prostie care a mai rămas în carte mi se datorează în întregime.

Nu aș fi putut scrie această carte fără susținerea de la Wellcome Trust (printr-un Engagement Fellowship), a Dr. Mortimer și Theresa Sackler Foundation și a Canadian Institute for Advanced Research (de unde conduc, în parteneriat, programul lor despre Creier, minte și conștiență). Le sunt recunoscător pentru susținerea pe care mi-au acordat-o, atât ca timp, cât și ca cercetare întreprinsă de grupul meu. Și mulțumesc, Universitatea din Sussex, pentru că mi-ai oferit un cămin academic atâția ani.

Scanned with OKEN Scanner

Mulțumirile finale sunt rezervate agentului meu literar și editorilor. Will Francis, de la Janklow & Nesbit, m-a încurajat în primul rând să scriu această carte, m-a ajutat să formulez propunerea, să găsesc editori excelenți și, în general, m-a păstorit pe parcursul întregului proiect, de la început până la sfârșit. Mulțumiri, la fel, excelenților editori, Rowan Cope și Laura Hassan de la Faber & Faber, Stephen Morrow de la Dutton/Penguin și copy-editorului Eleanor Rees.

Scanned with OKEN Scanner

Note

Prolog

meditație asupra mortalității: Barnes (2008).

1. Adevărata problemă

teoria spațiului de lucru global: Baars (1988); Dehaene & Changeux (2011); Mashour et al. (2020); Shanahan (2010).

gradul de flexibilitate comportamentală: Un alt mod de a gândi teoria spațiului de lucru global este ca o teorie a „conștienței accesului", în raport cu „conștiența fenomenală". Conștiența fenomenală se referă în mod explicit la experiență. Conștiența de acces pune accent pe funcțiile cognitive în detrimentul experienței. Când o stare mentală este „conștientă de acces", înseamnă că starea mentală este disponibilă pentru tot felul de funcții cognitive, inclusiv raționament, decizie și controlul comportamentului. Vezi Block (2005).

teorii ale gândirii de ordin superior: Există multe varietăți de teorii de ordin superior. Vezi R. Brown et al. (2019); Fleming (2020); Lau & Rosenthal (2011).

definiția unei gene s-a schimbat: Portin (2009).

Iată cum o descrie el: Chalmers (1995a), p. 201.

Chalmers vine cu o altă propunere: Mai recent, Chalmers a introdus „meta-problema" conștienței – de ce se crede că ar exista o problemă grea a conștienței. (Chalmers, 2018). Meta-problema este, de fapt, o versiune a problemei ușoare, deoarece se referă la explicarea comportamentului – în acest caz, comportamentul verbal al oamenilor care își manifestă convingerea despre o problemă grea a conștienței. Ceea ce îmi place la meta-problemă este că putem aprecia faptul că este o problemă și că o studiem, indiferent de atitudinea metafizică pe care o adoptăm cu privire la conștiența însăși.

Un „mecanism": Craver & Tabery (2017).

Chiar și când am explicat: Chalmers (1995a), p. 203, cursive în original.

ele pot fi considerate sinonime: Diferența dintre fizicalism și materialism este în mare măsură istorică. Materialismul este un cuvânt mai vechi decât fizicalismul. Unii susțin și că fizicalismul este o „teză lingvistică" (adică fiecare afirmație lingvistică este echivalentă cu o afirmație fizică), în timp ce materialismul este o afirmație mai generală despre natura lucrurilor. Vezi Stoljar (2017).

Scanned with OKEN Scanner

fizicalismul e de la sine înțeles: Deși majoritatea funcționaliștilor sunt fizicaliști, este posibil să fii funcționalist fără a fi fizicalist.

Creierii sunt foarte diferiți: Captivantul *The Idea of the Brain* (2020) al lui Matthew Cobb relatează istoria interpretării modului în care funcționează creierul pe baza tehnologiei dominante a zilei (și uneori invers).

Un computer care joacă Go: Silver et al. (2017). Povestea programului original, AlphaGo, este frumos spusă într-un film cu același nume: www.alphagomovie.com. Unii ar putea sugera că aceste programe sunt descrise cu mai multă acuratețe jucând, mai degrabă, „istoria lui Go", decât Go în sine.

există o întrebare valabilă: O versiune mai sofisticată a acestui argument a fost dezvoltată de John Searle în celebrul său experiment de gândire „Camera chinezească". Nu am folosit acest exemplu aici, deoarece argumentul lui Searle vizează mai degrabă inteligența (sau „înțelegerea") decât conștiența. (Searle, 1980).

o fundătură empirică: Filosoful John Perry a spus: „Dacă te gândești la conștiență suficient de mult, fie devii panpsihist, fie intri în administrație." Poate că problema constă mai degrabă în gândirea, decât în practica (științei). Pentru o susținere articulată a panpsihismului, recomand *Galileo's Error* de Philip Goff (2019). Citatul lui Perry apare într-un articol din 2018 din *Quartz* de Olivia Goldhill: qz.com/1184574/the-idea-that-everything-from-spoons-to-stones-are-conscious-is-gaining-academic-credibility.

misterianism: McGinn (1989).

niciun om nu ar putea înțelege vreodată: Pentru un argument împotriva ideii că unele lucruri vor fi întotdeauna dincolo de înțelegerea umană, vezi Deutsch (2012).

cel mai onest din punct de vedere intelectual: Mai există un „ism" care merită menționat, deși este un „ism" mai informal decât cele discutate în textul principal. Iluzionismul afirmă despre conștiența (fenomenală) că este o iluzie introspectivă – că atunci când introspectăm despre stările conștiente, le denaturăm conferindu-le proprietăți fenomenale – qualia – pe care, de fapt, nu le au. Într-un text – cu care nu sunt de acord – iluzionismul spune că stările conștiente nu există cu adevărat. În altul – cu care simpatizez mai mult – iluzionismul spune că experiențele conștiente există, dar nu sunt ceea ce credem noi că sunt. Este posibil, deși nu e clar, ca acest text al iluzionismului să fie compatibil cu ceea ce voi spune în această carte. Pentru mai multe despre iluzionism, vezi Frankish (2017).

Scanned with OKEN Scanner

orice experiență conștientă, oricare ar fi ea: Există cel puțin două tipuri de zombi filosofici. „Zombi comportamentali" nu se pot distinge de omologii lor conștienți din exterior – în termeni de comportament. „Zombi neurologici" vin cu fantezia suplimentară de a fi identici cu o creatură conștientă și din interior. Un zombi neurologic are aceeași structură internă și poate fi, de asemenea, făcut din aceeași wetware (echivalentul biologic al software, n. trad.) electrochimic, ca și creatura conștientă a cărei versiune este zombi. Conștiența lipsește cu desăvârșire la toate soiurile de zombi.

o rețea vastă: Gidon et al. (2020); Herculano-Houzel (2009).

imaginați-vă inimaginabilul: apărătorii zombi ar putea răspunde că ceea ce contează este doar posibilitatea lor de a avea logică, nu imaginabilitatea lor, dacă luăm în considerare legile fizicii din acest univers anume. Nu sunt de acord. Un A380 care zboară în marșarier poate fi logic posibil, dacă luăm în considerare principiile alternative ale aerodinamicii, dar a accepta o posibilitate nu elucidează modul în care zboară cu adevărat un A380, în această lume, cu aceste legi ale fizicii și aerodinamicii. Ceea ce vreau să știu este cum anume un creier real (și corp și așa mai departe), respectând legile fizicii din acest univers, modelează și dă naștere unei experiențe conștiente în acest univers.

informații de la mai mulți: modul în care „problema reală" abordează lucrurile nu este nou – cel puțin nu în totalitate. Chalmers însuși a descris strategii similare sub forma „problemei de cartografiere" (Chalmers, 1996) și a „principiului coerenței structurale" (Chalmers, 1995a), care fac parte din descrierea sa inițială a problemei grele. Există, de asemenea, o acumulare cantitativă și foarte influentă de lucrări în „neurofenomenologie", care încearcă să potrivească proprietățile fenomenologice cu aspecte ale creierului și ale activității acestuia (Thompson, 2014; Varela, 1996). Există, totuși, diferențe de accent între aceste poziții (Seth, 2009, 2016b).

Nu s-a scris nimic demn de citit: Sutherland (1989).

devin metoda dominantă: Crick & Koch (1990). Aproximativ în aceeași perioadă, filozoful american Daniel C. Dennett și-a publicat influenta *Consciousness Explained* (1991). Citirea acestei cărți la începutul anilor 1990 a fost, pentru mine, un punct de cotitură. Rămâne o lectură dificilă și lămuritoare. Vezi LeDoux et al. (2020) și Seth (2017, 2018) pentru mai multe despre istoria științei conștienței.

mecanisme neuronale minime: Crick & Koch (1990).

NCC pentru acea experiență particulară: rețineți că NCC-urile sunt

Scanned with OKEN Scanner

interpretate, de obicei, ca fiind în relație cu anumite regiuni ale creierului, dar nu este obligatoriu să fie așa. Definiția unui NCC se referă la un mecanism neuronal care ar putea fi implementat într-o serie de regiuni ale creierului. Pentru un anumit NCC, circuitele cerebrale implicate chiar se pot schimba în timp (G. M. Edelman & Gally, 2001).

Activitate cerebrală în desfășurare: O subtilitate aici este că regiunile creierului care apar în studiile de rivalitate binoculară sunt adesea asociate cu tranziții între percepțiile conștiente, ceea ce este important, deoarece percepția schimbării nu este același lucru cu schimbarea percepției. Vezi Blake et al. (2014). Această diferență este reluată în capitolul 6.

este dificil: se vorbește despre aceste probleme de ani de zile dar, în general, au fost fie respinse ca irelevante, fie băgate sub preș în alt mod. Două articole din 2012 au formulat, în sfârșit, problema: Aru et al. (2012); de Graaf et al. (2012).

nu trebuie confundat: Unele experimente au făcut încercări curajoase de a diferenția percepția conștientă de atenție și de răspunsul comportamental. Rezultatele așa-numitelor paradigme „fără răspuns", în care voluntarii nu răspund comportamental în legătură cu ceea ce percep, sunt deosebit de interesante. În multe dintre aceste studii, NCC rămasă nu include regiunile frontale ale creierului, care sunt esențiale pentru teorii precum cea a spațiului de lucru global și cea a gândirii de ordin superior. Vezi Frässle et al. (2014); Tsuchiya et al. (2015); și Raccah et al. (2021) pentru o discuție recentă.

știință pe deplin satisfăcătoare: Un alt mod de a îndeplini această ambiție vine de la filosofii Susan Hurley și Alva Noë, care fac diferența între lacune explicative „comparative" referitoare la faptului că diferitele experiențe au proprietățile fenomenologice specifice pe care le au, precum și la decalajul explicativ „absolut", care este problema (grea) referitoare la de ce și cum există ceva precum fenomenologia. Ne putem gândi la problema reală ca fiind abordare exhaustivă a lacunelor explicative comparative pentru a rezolva și, poate, dizolva decalajul explicativ absolut. Vezi Hurley & Noë (2003).

Gândire vitalistă: Vitalismul susține că „organismele vii sunt fundamental diferite de entitățile nevii, deoarece conțin un element non-fizic sau sunt guvernate de principii diferite comparativ cu lucrurile neînsuflețite" (Bechtel & Williamson, 1998). Chiar și astăzi, majoritatea copiilor de vârstă preșcolară tind să prefere explicațiile vitaliste ale vieții explicațiilor mai moderne (Inagaki & Hatano, 2004). Paralelele istorice dintre vitalism și știința conștienței

Scanned with OKEN Scanner

au fost explorate deosebit de intens de filosoafa Patricia Churchland (Churchland, 1996).

Un evrica zgomotos: nevoia de o soluție de tip evrica poate explica parțial atracția persistentă pentru teoriile conștienței bazate pe mecanica cuantică, dintre care cele mai multe se leagă de cartea matematicianului Roger Penrose, *The Emperor's New Mind,* publicată în 1989. Deși nu se poate nega ideea că o teorie viitoare bazată pe cuantică să aibă ceva util de spus despre conștiență, încercările de până acum mi se pare că expun un silogism fals: mecanica cuantică este misterioasă, conștiența este misterioasă, de aceea trebuie că există o legătură între ele.

2. Măsurarea conștienței

Joachim Dalencé: www.encyclopedia.com/science/dictionaries-thesauruses-pictures-and-press-releases/dalence-joachim.

Inventarea temperaturii: H. Chang (2004).

Ar putea funcționa aceeași abordare: probabil prima persoană care a pus această întrebare și a și răspuns a fost Thomas Henry Huxley („Bulldogul lui Darwin"). Într-o prelegere din 1870, el a spus: „Cred că, mai devreme sau mai târziu, vom ajunge la un echivalent mecanic al conștienței, așa cum am ajuns la un echivalent mecanic al căldurii". Citat în Cobb (2020), p. 113.

potențialul de a transforma înțelegerea științifică: Seth et al. (2008).

peste patru milioane: Weiser et al. (2008).

monitorul pentru „indice bispectral": Myles et al. (2004).

a rămas controversată: Nasraway et al. (2002).

de patru ori mai mulți neuroni: Herculano-Houzel (2009).

nici un motiv de îndoială: vezi Lemon & Edgley (2010). Cerebelul apare adesea ca o „zonă de interes" în studiile de neuroimagistică. Însă, aceste descoperiri nu prea sunt discutate, deoarece cerebelul nu a fost centrul experimentului, așa că cercetătorii nu știu ce să spună despre el.

diferențe în consumul de energie al creierului: DiNuzzo & Nedergaard (2017).

complexitatea acestor tipare: Ferrarelli et al. (2010); Massimini et al. (2005).

reprezentare comprimată: Pentru a crea o reprezentare comprimată a unei secvențe doar de 1 (sau de 0), în cazul unei secvențe de lungime n, tot ce trebuie să faci este să specifici primul 1 (sau 0), apoi să specifici n repetări. La cealaltă extremă, o secvență complet aleatorie

Scanned with OKEN Scanner

nu poate fi comprimată deloc: trebuie să specifici fiecare 1 și 0 în locul său corespunzător pentru a o regenera cu acuratețe. Secvențele care conțin o anumită structură previzibilă vor avea reprezentări comprimate undeva la mijloc. Lungimea reprezentării comprimate optim se numește complexitatea Kolmogorov-Chaitin-Solomonoff. Complexitatea LZW aproximează și oferă o limită superioară a acestei cantități.

studiu reper: Casali et al. (2013).

serie de studii: Schartner et al. (2017b); Schartner et al. (2015). Ne-am asigurat, în aceste studii, că schimbările în complexitatea LZW nu sunt reflectări banale ale altor – mai cunoscute – modificări ale semnalelor EEG ale creierului care însoțesc pierderea conștienței, cum ar fi creșterea puterii undelor „delta" de joasă frecvență, despre care știm că au loc în timpul somnului.

același model de rezultate: Casarotto et al. (2016).

abordări clinice: Vezi, de exemplu, scara Glasgow a comei, utilizată pe scară largă (Teasdale & Murray, 2000).

Această afecțiune rară: sindromul locked-in este cel mai adesea cauzat de leziuni ale punții – o parte a trunchiului cerebral care conține căi neuronale importante ce leagă măduva spinării cu alte părți ale creierului.

a scris o carte întreagă: Bauby (1997). S-ar putea să credeți că a fi locked-in este o soartă mai rea decât moartea, dar în timp ce unii oameni aflați în această stare consideră că viața este intolerabilă, o proporție surprinzătoare raportează o calitate rezonabilă a vieții (Rousseau et al., 2015). Iată încă o dovadă că e periculos să facem presupuneri despre cum anume decurg lucrurile în interior, luând-ne după aparențele din exterior.

un caz ciudat: Povestea este relatată și aici: www.humanbrainproject.eu/en/follow-hbp/news/measures-of-consciousness-and-a-tale-of-cultural-loss.

Owen a concluzionat: Owen et al. (2006). Adrian Owen relatează povestea descoperirii echipei sale și implicațiile ei pentru medicină și cercetarea conștienței, în cartea sa *Into the Grey Zone* (Owen, 2017).

mod laborios de comunicare: Monti et al. (2010). Având în vedere că pacienții pot răspunde doar „da" sau „nu" în această configurație, cercetătorii trebuie să fie foarte atenți atunci când își selectează întrebările. A întreba dacă cineva suferă este rezonabil, dar cum e să întrebi dacă vrea să rămână în viață? Există multe dileme etice în acest sens. În 2016, scriitoarea Linda Marshall Griffiths, producătorul Nadia Molinari și cu mine am creat o dramă audio,

Scanned with OKEN Scanner

The Sky Is Wider, care explorează aceste probleme. https://www.lindamarshallgriffiths.co.uk/the-sky-is-wider-best-single-drama.

o analiză recentă: Naci et al. (2017).

poate fi limitat: Un exemplu mai cunoscut al acestei probleme a fost încercarea de a defini și măsura inteligența în raport cu o singură scară, cum ar fi „IQ". Acest proiect s-a clătinat întotdeauna în fața varietății enorme de comportament inteligent observat între indivizi, culturi și specii. Vezi Bayne et al. (2016) pentru o abordare multidimensională a caracterizării nivelului conștient.

comunici cu oamenii în timpul viselor lucide: Vezi Konkoly et al. (2021).

Puțin câte puțin: Albert Hofmann, „LSD – My Problem Child", http://psychedelic-library.org/child1.htm.

o generație pierdută: vezi *How to Change Your Mind* a lui Michael Pollan, pentru mai multe informații despre această istorie (Pollan, 2018).

modificări izbitoare ale dinamicii creierului: Carhart-Harris et al. (2012).

rezoluție temporală înaltă: datele MEG și EEG au, de obicei, o rezoluție temporală foarte mare, de ordinul milisecundelor, deoarece reflectă direct activitatea electrică a populațiilor de neuroni. RMNf este foarte lent, prin comparație, cu o scală temporală naturală de secunde. Aceasta din cauză că scannerul RMN obișnuit funcționează lent, permițând, de obicei, o măsurătoare la fiecare secundă sau la două secunde, dar și din cauza scalei temporale joase a semnalului de oxigenare a sângelui pe care îl măsoară RMNf.

rezultatele au fost clare: Schartner et al. (2017a). Un alt studiu mai recent a găsit același model de rezultate cu DMT (Timmermann et al., 2019).

Descoperirile noastre: Într-un studiu ulterior, condus de Lionel Barnett, am descoperit că starea psihedelică implică reduceri semnificative ale „fluxului de informații" dintre regiunile cortexului. Acest lucru vorbește din nou despre o pierdere a structurii perceptive în starea psihedelică. Vezi Barnett et al. (2020).

O lucrare publicată în 1998: Tononi & Edelman (1998). Edelman a murit în 2014 – am scris unul dintre multele necrologuri care aduceau un omagiu acestui remarcabil om de știință (Seth, 2014a).

par legați împreună: Unii filosofi au pus sub semnul întrebării presupunerea că experiențele conștiente sunt în mod necesar unificate (Bayne, 2010). Un scenariu în care se discută adesea posibilitatea unei conștiențe neunificate apare la pacienții cu creier divizat, cărora li s-au separat chirurgical emisferele corticale. Mai multe despre asta în capitolul 3.

complexitatea neuronală: Tononi et al. (1994).

densitate cauzală: Seth et al. (2011a); Seth et al. (2006).

Scanned with OKEN Scanner

Rafinarea matematicii: Pentru o comparație a măsurilor complexității în contextul neuroștiinței, a se vedea Seth et al. (2011a) și Mediano et al. (2019).
Alte măsuri: Demertzi et al. (2019); Luppi et al. (2019).

3. Phi

IIT spune: Unele dintre principalele articole despre IIT din grupul lui Tononi sunt: Tononi (2008); Tononi (2012); Tononi et al. (2016). Pentru un manifest accesibil, a se vedea Koch (2019).
Lucrarea mea recentă: Seth et al. (2006).
un pas gigantic: www.scientificamerican.com/article/is-consciousness-universal.
intuiție periculoasă: vezi, de exemplu, criticile lui Scott Aaronson: www.scottaaronson.com/blog/?p=1799.
revendicarea principală a teoriei: Φ poate fi, de asemenea, gândit ca o modalitate de a măsura proprietatea „emergenței". Emergența este un concept foarte general, care se referă la modul în care proprietățile macroscopice (cum ar fi un stol) apar din, sau se leagă de componentele lor microscopice (păsările individuale). Vezi Hoel et al. (2013); Rosas et al. (2020); Seth (2010).
expresia supremă a viziunii bazate pe temperatură: Tononi și Koch folosesc ei înșiși această analogie (Tononi și Koch, 2015).
moduri în care un sistem poate eșua: Acestea sunt adaptate după Tononi (2008).
funcționează la fel de bine: aceasta este o primă aproximare, lăsând deoparte lucruri precum reglarea contrastului care ar putea funcționa în întreaga matrice.
intervenție chirurgicală de divizare: operațiile cu creier divizat implică tăierea mănunchiului masiv de fibre nervoase care interconectează emisferele corticale, corpul calos. Intervenția chirurgicală poate avea succes în ameliorarea epilepsiei severe, dar este rar efectuată acum pentru că alte opțiuni de tratament, mai puțin invazive, sunt din ce în ce mai disponibile. Pacienții cu creier divizat păstrează un anumit grad de conectivitate între emisfere dar, în scopul urmărit de acest exemplu, să ne imaginăm că întregul lor creier este tăiat în jumătate. Vezi de Haan et al. (2020).
IIT răspunde clar: Tononi et al. (2016).
IIT propune axiome: Pe lângă axiomele integrării și informației, IIT propune alte trei: conștiența există, este compusă din multe elemente și este exclusivă la o anumită scală spațio-temporală (Tononi et

Scanned with OKEN Scanner

al., 2016). Filosoful Tim Bayne a criticat IIT pe baza faptului că axiomele propuse de acesta, în special cea finală privind „excluderea", ar putea să nu fie în mod evident adevărate (Bayne, 2018).

Măsoară reducerea incertitudinii: în teoria informației, informația – reducerea incertitudinii – este măsurată folosind o cantitate numită entropie. Entropia (denumită de obicei S) este o funcție a numărului de stări distincte în care se poate afla un sistem, împreună cu probabilitatea de a fi în fiecare dintre aceste stări. Este dată de ecuația $S = -\Sigma p_k \log(p_k)$. În cuvinte, aceasta înseamnă că, pentru fiecare stare (k) a unui sistem, înmulțim probabilitatea de a fi în acea stare cu logaritmul acelei probabilități, apoi însumăm aceste valori pentru toate stările. Pentru orice sistem dat, valoarea entropiei este maximă atunci când fiecare stare este la fel de probabilă. Un zar ușor are o entropie de aproximativ 2,5 biți (când logaritmul e în baza 2). Un zar greu va avea o entropie mai mică.

A ști totul despre un mecanism: valoarea Φ a unui sistem este, prin urmare, mai mult o afirmație despre mecanismul său (cum anume este conectat) decât despre dinamica sa (ce anume face). Într-adevăr, versiunile recente ale IIT descriu Φ în termeni de „putere ireductibilă cauză-efect", care este o afirmație despre mecanism, nu despre dinamică (Tononi et al., 2016).

împărțirea sistemului: din punct de vedere tehnic, această diviziune se numește partiția minimă de informații. Există, de asemenea, probleme complicate cu privire la modul cel mai bun de a trata partițiile de diferite dimensiuni, deoarece o partiție mai mare va fi capabilă, în virtutea faptului că are mai multe elemente, să genereze mai multe informații.

Ar putea o țară întreagă: Potrivit IIT, o „țară conștientă" ar putea apărea dacă Φ este maximizat la o scală spațio-temporală care operează pe o țară întreagă, poate cu indivizii analogi neuronilor unui creier. O astfel de situație ar avea implicația particulară că, odată ce o țară este conștientă, elementele sale individuale – oamenii – nu ar mai fi conștiente individual. Acest scenariu bizar a fost explorat de filozoful american Eric Schwitzgebel: schwitzsplinters.blogspot.com/2012/03/why-tononi-should-think-that-united.html.

a lumina creierul în interior: Deisseroth (2015).

să se facă într-adevăr: vezi www.templetonworldcharity.org/accelerating-research-consciousness-our-structured-adversarial-collaboration-projects. Experimentul „inactiv" versus „inactivat" a fost propus de Umberto Olcese și Giulio Tononi.

abordarea „it din bit": Wheeler (1989).

Scanned with OKEN Scanner

diversele noastre versiuni: Barrett & Seth (2011); Mediano et al. (2019). Din punct de vedere matematic, măsurile noastre funcționează mai degrabă cu distribuția empirică a unui sistem, decât cu distribuția entropiei sale maxime.

cum se dezvoltă: o propunere interesantă este aceea că „spațialitatea" omniprezentă a experienței vizuale este explicată de anatomia asemănătoare unei grile, care se găsește la nivelurile inferioare ale cortexului vizual (Haun & Tononi, 2019).

4. Percepția în sens invers

semnale de la fiecare nivel: Ierarhiile perceptuale nu sunt strict izolate unele de altele. O regulă generală este aceea că, cu cât te îndepărtezi de periferia senzorială, cu atât ai mai multă interacțiune cross-modală. Vezi Felleman & Van Essen (1991); Stein & Meredith (1993).

Experimente mai recente: Grill-Spector & Malach (2004).

teoria computațională a vederii: Marr (1982).

niveluri de performanță impresionante: He et al. (2016).

schimb încântător: Anscombe (1959), p. 151, cursiv în original.

judecată și inferență: Din *Optica lui Ibn al-Haytham,* c.1030, tradusă în Sabra (1989). Istoria este spusă mai detaliat în minunatul *The Predictive Mind* (2013) al lui Jakob Hohwy.

Helmholtz s-a văzut pe sine: Swanson (2016).

cei din familii înstărite nu: Bruner & Goodman (1947).

Conținut perceptiv: Gregory (1980).

teoriile diferă în detalii: Clark (2013); Clark (2016); Howwy (2013); Rao & Ballard (1999).

predicții top-down/de sus în jos: nu toate predicțiile sunt de sus în jos și nici erorile de predicție nu sunt neapărat de jos în sus. „Predicțiile bottom-up/de jos în sus" pot fi considerate constrângeri care reflectă aspecte globale și stabile ale inferenței perceptive (Teufel & Fletcher, 2020). Un exemplu de constrângere perceptivă ar putea fi supra-reprezentarea orientărilor verticale și orizontale în imaginile naturale. Influența umbrelor asupra luminozității – discutată mai târziu – ar putea fi, de asemenea, o constrângere predictivă bottom-up.

Dennett subliniază: Vezi Dennett (1998). Iată un alt exemplu: pentru a auzi muzică, nu avem nevoie de o trupă mică în cap și de un sistem complex de microfoane intracraniene. Tot Daniel Dennett este autorul jocului de cuvinte „figment mental".

Scanned with OKEN Scanner

culoarea este locul: Gasquet (1991). Acest citat a fost atribuit și lui Paul Klee.

Să agreăm în legătură cu halucinațiile noastre: Mulțumim mult artistului rap Baba Brinkman pentru această replică provocatoare despre realitate. Mai multe despre percepția a „ceea ce este real", în capitolul 6.

Cărțulie despre iluziile vizuale: Un studiu recent al Biyu He și colegilor săi a analizat modul în care dinamica neuronală se schimbă atunci când o imagine Mooney este recunoscută, în comparație cu atunci când nu este. Vezi Flounders et al. (2019).

acestea și... alte ipoteze: Brainard & Hurlbert (2015); Witzel et al. (2017).

schimbă ceea ce vezi în mod conștient: Eye Benders (Gifford & Seth, 2013).

exemple auditive convingătoare: Chris Darwin are câteva exemple excelente de vorbire sinusoidală online la www.lifesci.sussex.ac.uk/home/Chris_Darwin/SWS. Eu folosesc un alt exemplu în discuția mea TED din 2017: www.ted.com/talks/anil_seth_your_brain_hallucinates_your_conscious_reality. Există, de asemenea, echivalente auditive ale Rochiei. Un exemplu este un sunet pe care unii oameni îl aud ca „Yanny", iar alții ca „Laurel" (Pressnitzer et al., 2018). În 2020, a apărut un videoclip TikTok în care un zgomot ambiguu de la o jucărie ieftină poate fi auzit fie ca „green needle", fie ca „brainstorm", în funcție de cuvintele pe care le citiți (time.com/5873627/green-needle-brainstorm- explicat).

se construiește experiența perceptivă: Vezi de Lange et al. (2018) pentru o sinteză a experimentelor care arată modul în care așteptările modelează percepția.

Percepție exactă – „veridică": un alt stimul din zona intuitivă pentru a te gândi la veridicitatea perceptivă este vederea periferică – acea zonă a câmpului vizual situată departe de centrul privirii (vederea foveală). Periferia vizuală are o densitate mult mai mică a fotoreceptorilor decât foveea și, totuși, experiența vizuală la periferie nu pare neclară. Înseamnă aceasta că vederea periferică este mai puțin veridică decât vederea foveală, deoarece manifestă o „iluzie de claritate" (sau, mai precis, o iluzie de „neîncețoșare")? Nu! Claritatea și neclaritatea sunt proprietăți ale experienței perceptive relative la datele senzoriale pentru care este reglată fiecare parte a sistemului vizual. Vezi Haun (2021) pentru o discuție revelatoare, Lettvin (1976) pentru contextul istoric și Hoffman et al. (2015) pentru o discuție amplă despre (non-)veridicitate perceptivă.

Locke a propus: John Locke, *An Essay concerning Human Understanding* (1689), a 14-a ediție (1753). Deși exemplul culorii este intuitiv,

Scanned with OKEN Scanner

filosofii au dezbătut mult dacă aceasta contează într-adevăr ca o calitate secundară (Byrne & Hilbert, 2011). O diferențiere similară, care apare în literatura filosofică, se face între diferitele tipuri de „tipuri" – unde diferențele constau în condițiile necesare pentru ca ceva să existe. De exemplu, banii necesită convenții sociale pentru a exista și, astfel, sunt de „tip social". Apa nu necesită convenții sociale pentru a exista, așa că este de „tip natural".

5. Vrăjitorul din Odds

Vrăjitorul din Odds: Mulțumim din nou lui Baba Brinkman, care a inspirat acest titlu. Vine din al său *Rap Guide to Consciousness* (2018), pentru care am fost consilier științific. Vezi https://bababrinkman.com/shows/#consciousness.

pajiștea este udă: Acest exemplu este adaptat după F. V. Jensen (2000). Pentru o fundamentare amănunțită a temei raționamentului abductiv, nu puteți găsi o carte mai bună decât *Inference to the Best Explanation* (2004) a lui Peter Lipton.

Regula în sine: Iată cum este scrisă în mod normal regula lui Bayes:

$$p(H|D) = \frac{p(D|H) * p(H)}{p(D)}$$

$p(H|D)$ este posteriorul – probabilitatea ipotezei H, având datele D, $p(D|H)$ este probabilitatea – probabilitatea datelor furnizate în contextul ipotezei, iar $p(H)$ și $p(D)$ sunt probabilitățile precedente ale ipotezei și, respectiv, ale datelor. De reținut că găsirea unui număr pentru $p(D)$ poate fi dificilă dar, din fericire, nici nu prea e necesar. Atunci când scopul este de a găsi cel mai probabil posterior dintr-o serie de alternative, așa cum se întâmplă adesea, $p(D)$-urile se anulează frumos.

trageți concluzia greșită că aveți o boală urâtă: în 2009, guvernul SUA a fost împotriva mamografiei în masă pentru identificarea cancerului de sân la femeile de patruzeci de ani exact din acest motiv. La acea vreme, mamografiile aveau sensibilitate de aproximativ 80%, ceea ce înseamnă că au detectat aproximativ 80% dintre femeile din această grupă de vârstă care aveau cancer de sân. Dar testele au semnalat cancer și la aproximativ 10% dintre femei care nu aveau boala la acel moment. Cancerul de sân are incidență mică – sau „rată de bază" – de circa 0.04 % în acest grup de vârstă. Dacă aplicăm regula lui Bayes pornind de la această rată de bază ca precedentă, putem calcula că probabilitatea de a avea cancer de sân la o mamografie pozitivă este de numai aproximativ 3%. Din fiecare 100 de femei

Scanned with OKEN Scanner

testate pozitiv, 97 nu au cancer, dar vor deveni anxioase și vor face investigații suplimentare potențial costisitoare și invazive. Morala acestei povești este că e nevoie de îmbunătățirea sensibilității și specificității testării, iar mamografiile sunt, într-adevăr, mult mai bune acum. Un studiu recent din Marea Britanie a sugerat că screeningul femeilor în vârstă de patruzeci de ani ar putea fi acum util. Vezi McGrayne (2012); Duffy et al. (2020).

S-a aplicat inferența bayesiană: istoria surprinzător de controversată a analizei bayesiene este descrisă superb în *The Theory That Would Not Die* (2012) a lui Sharon McGrayne.

În filosofia științei: Pentru mai multe despre subiect, vezi Lakatos (1978) și Seth (2015b).

a-mi răsturna convingerea Bayesiană: îmi ridic pălăria în fața lui Paul Fletcher și Chris Frith, care au spus același lucru despre teoria lor bayesiană a halucinațiilor și a iluziei în schizofrenie (Fletcher & Frith, 2009).

trebuie să se întâmple ceva: strict, X este o variabilă aleatorie, deoarece valoarea sa este determinată de o distribuție de probabilitate. Exemplul din figura de la p. 106 este o distribuție continuă de probabilitate (numită și funcție de densitate de probabilitate), deoarece X poate lua orice valoare în intervalul permis. Dacă X ar putea lua doar valori specifice – de exemplu „capete" sau „cozi" – am avea o distribuție de probabilitate discretă.

încrederea pe care o are creierul în această estimare a probabilității: Există două moduri de a interpreta mapările între convingerile bayesiene și creier. Interpretarea mai superficială este că noi, ca observatori externi, folosim aceste convingeri pentru a ne reprezenta lucruri, la fel cum am putea folosi o hartă fizică pentru a ne reprezenta împrejurimile. Din această perspectivă, activitatea neuronală observată ar putea reprezenta anumite stări de lucruri pentru noi, ca oameni de știință. Interpretarea mai bună este că creierul folosește aceste convingeri (sau ceva asemănător) pentru a reprezenta lucrurile pentru el însuși. Această a doua interpretare exemplifică ipoteza „creierului bayesian" și este fundamentală pentru ideea de creier ca mașină de predicție. Incapacitatea de diferențiere între aceste două semnificații ale „reprezentării" a fost sursa multor confuzii în întreaga știință cognitivă și neuroștiință. Vezi Harvey (2008).

toate predicțiile de top-down: Unii se referă la căile bottom-up și top-down ca „feedforward" și, respectiv, „feedback". Din perspectiva procesării predictive, aceasta este o cale eronată. În inginerie, „feedback" este, de obicei, asociat cu un semnal de eroare, utilizat

Scanned with OKEN Scanner

apoi pentru a regla semnalul de control „feedforward". Prin urmare, în procesarea predictivă, conexiunile bottom-up sunt, acum, conexiuni de „feedback", deoarece aceste conexiuni transmit semnalele de eroare. Complicand lucrurile și mai mult, așa cum am menționat mai devreme, unele predicții (globale, stabile) pot fi încorporate în semnale bottom-up, conexiunile top-down transmitând acum erori de predicție (Teufel & Fletcher, 2020).

procesare predictivă: folosesc procesarea predictivă în această carte ca prescurtare pentru o varietate de teorii care includ, dar nu se limitează la codificare predictivă, care încorporează mecanismul de bază al minimizării erorilor de predicție. Procedând astfel, nu vreau să diminuez diferențele dintre ele, care sunt și interesante și importante (Hohwy, 2020a).

încearcă să minimizeze aceste erori: În termeni bayesieni, un model generativ este specificat printr-o combinație de o precedentă și o probabilitate. Poate fi gândită ca „probabilitatea comună a ipotezei și a datelor luate împreună", *p(H, D)*. Privind astfel lucrurile, aparatul matemtic susține afirmația că minimizarea erorilor de predicție aproximează inferența bayesiană (Buckley et al., 2017).

cascadă în continuare: minimizarea erorilor de predicție ierarhică sugerează că părțile creierului implicate în percepție ar trebui să fie bogat înzestrate cu conectivitate top-down, care transmite semnale de la niveluri ierarhice superioare la cele inferioare. Multe studii au arătat că chiar așa e; vezi, de exemplu, Markov et al. (2014). Prezența unei astfel de conectivități bogate top-down este mai greu de explicat dintr-o perspectivă bottom-up a percepției.

ca de nicăieri: Vezi Kuhn et al. (2008) pentru o revizuire a psihologiei și magiei.

ponderare/cântărire de precizie: Vezi Feldman & Friston (2010). Ponderarea preciziei se realizează prin modificarea precedentelor despre precizii – așa-numitele hiperprecedente – astfel încât precizia inferată să fie crescută sau micșorată.

demonstrație video: Vezi și Simons & Chabris (1999).

pentru a acționa eficient: colegul meu de la Universitatea Sussex, Andy Clark, a susținut de mult timp formulări „orientate spre acțiune" de procesare predictivă. Cartea sa din 2016, *Surfing Uncertainty* (2016) este o resursă-reper.

Generarea de acțiuni: așa cum susține cercetătorul în neuroștiință György Buzsáki în cartea sa *The Brain from Inside Out*, această perspectivă vine cu provocări și deschide noi oportunități pentru neuroștiința experimentală. Majoritatea experimentatorilor, în niciun

Scanned with OKEN Scanner

caz toți, studiază creierul mai degrabă examinându-i activitatea ca răspuns la stimularea externă, decât ca pe un sistem activ intrinsec dinamic. Vezi Buzsáki (2019), precum și Brembs (2020).

inferență activă: Friston et al. (2010)

Cercul virtuos: Alexander Tschantz, Christopher Buckley, Beren Millidge și cu mine am dezvoltat noi algoritmi machine-learning pe această bază, care sunt capabili să învețe modele generative din cantități mici de date (Tschantz et al., 2020a). Interesant este faptul că posibilitatea unor predicții „bottom-up" (Teufel & Fletcher, 2020) are o aplicație promițătoare în acest context. Poate fi legată de tehnica de „amortizare" în machine-learning, în care un posterior bayesian aproximativ este calculat cu ajutorul unei treceri feedforward (bottom-up) printr-o rețea neuronală artificială antrenată corespunzător.

neatenție: Această idee poate fi testată experimental și, într-adevăr, se dovedește că în timpul acțiunii, sensibilitatea senzorială proprioceptivă este redusă, așa cum ar fi de așteptat (C. E. Palmer et al., 2016). Acțiunea care însoțește atenuarea senzorială oferă, de asemenea, o explicație clară a motivului pentru care nu ne putem gâdila pe noi înșine (H. Brown et al., 2013).

6. Contribuția privitorului

Era explorării interioare: Kandel (2012).

popularizat ulterior: Gombrich (1961)

contribuția celui care percepe: Seth (2019b).

După cum a spus Kandel: Kandel (2012), p. 204.

Brumă la Ennery: Acest exemplu este preluat din Albright (2012), adaptat în Seth (2019b). Vezi www.wikiart.org/en/camille-pissarro/hoarfrost-1873.

răzuieli-de-paletă: Expresia apare în recenzia satirică a artei impresioniste a lui Leroy, publicată în *Le Charivari*, la 25 aprilie 1874 – o recenzie care a creat termenul de impresionism.

Privire inocentă: Gombrich (1961).

picturile devin experimente: în ultimii săi ani, Pissarro a suferit de probleme serioase la ochi. La fel și colegii săi impresioniști, Claude Monet și Edgar Degas. Am putea să speculăm și să ne gândim dacă și cum anume afectarea vederii lor ar fi putut contribui la aprofundarea perspectivei lor artistice. În timp ce o anumită influență este plauzibilă – poate că au devenit mai sensibili la tiparele de lumină, decât la detaliile obiectelor – o abordare asemănătoare a

Scanned with OKEN Scanner

operei altui pictor, El Greco (1541–1614) vine să ne avertizeze în acest sens. Lucrările lui El Greco au inclus adesea figuri nenatural de alungite, o trăsătură care a fost atribuită, de unii, astigmatismului său pronunțat. Adică a pictat ce a văzut. Psihologii au atras atenția, însă, că, urmând această logică, El Greco și-ar fi văzut și pânza alungită, anulând orice efect al astigmatismului. Acest tip de eroare logică a fost numită „eroarea El Greco" și e încă interesantă pentru cercetătorii de percepție. Vezi Firestone (2013).

Când spunem: Gombrich (1961), p. 170.

natura experienței: revelația lui Gombrich a fost preluată, mai târziu, de scriitorul, criticul și artistul John Berger, a cărui carte din 1972, *Ways of Seeing*, începe cu textul „Relația dintre ceea ce vedem și ceea ce știm nu este niciodată clară". Spre deosebire de Gombrich, care era un conservator din punct de vedere cultural, Berger a subliniat influențele politice și culturale asupra percepției, accentuând faptul că oamenii și grupurile de oameni pot vedea diferit aceleași lucruri (Berger, 1972).

O predicție experimentală: de fapt, această predicție nu este chiar atât de simplă; vezi Press et al. (2020).

Suprimarea succesiunii continui de instantanee: Aceasta este o variantă a mai cunoscutei metode de „rivalitate binoculară", pe care am descris-o în capitolul 1 (Blake et al., 2014).

experimentul nostru: Pinto et al. (2015). Studiul nostru a fost mai complicat decât am rezumat aici. Am făcut un număr mare de studii de control pentru a exclude, cât de bine am putut, alți factori, cum ar fi prejudecățile care domină răspunsurile subiecților sau modul în care își concentrează atenția, pentru ca rezultatele să fie valabile. Vezi de Lange et al. (2018) pentru o sinteză a altor studii similare, Melloni et al. (2011) pentru o contribuție timpurie de valoare la această literatură.

tehnică eficientă: „Citirea creierului" implică antrenarea algoritmilor de machine-learning pentru a cataloga activitatea creierului în diferite categorii. Vezi Heilbron et al. (2020).

vezi fețe în lucruri: www.boredpanda.com/objects-with-faces.

o „mașină de halucinat": Suzuki et al. (2017).

Rețele ca aceasta: în mod specific, rețelele sunt rețele neuronale convoluționale profunde (deep convolutional neural networks = DCNN), care pot fi antrenate folosind algoritmi standard de propagare inversă. Vezi Richards et al. (2019).

inversează procedura: în modul standard „forward", o imagine este prezentată rețelei, activitatea este propagată în sus prin straturi,

Scanned with OKEN Scanner

iar output-ul rețelei ne spune ce „crede" că este în imagine. În algoritmul deep dream/vis profund – și în adaptarea lui Keisuke – acest proces este inversat. Fixăm output-ul rețelei și ajustăm input-ul, până când rețeaua ajunge într-o stare stabilă. Vezi Suzuki et al. (2017) pentru detalii.

mult mai convingătoare: experiența completă cu mașina de halucinat – în care o înregistrare panoramică este vizionată printr-un dispaly montat pe cap – este considerabil mai imersiv decât orice imagine statică. Exemplu de filmare aici: www.youtube.com/watch?v=TlMBnCrZZYY.

sunt așa cum sunt: atunci când modelele computaționale confirmă ipoteze despre circuitele neuronale, această abordare poate fi numită „neurofenomenologie computațională" – o versiune a neurofenomenologiei îmbunătățită din punct de vedere computațional, care poate fi urmărită în timp până la Francisco Varela (1996). În același spirit, Keisuke Suzuki, David Schwartzman și cu mine dezvoltăm noi versiuni ale mașinii de halucinat, care încorporează în mod explicit modele generative și, prin urmare, se apropie de ceea ce credem că se întâmplă în creierul real. Noile noastre aparate de halucinat sunt capabile să surprindă o varietate mult mai mare de experiențe halucinatorii decât originalul.

explorarea principiilor obiectității: Seth (2019b).

pictorul investighează: Merleau-Ponty (1964).

teoria contingenței senzorio-motorii: Vezi O'Regan (2011); O'Regan și Noë (2001). Ca orice altă teorie, aceasta se bazează pe o mulțime de lucrări anterioare, în acest caz, în special pe ideile lui James Gibson despre modul în care percepția depinde de acțiunea corporalizată (embodied action), precum și pe fenomenologia filozofică a lui Edmund Husserl și Maurice Merleau-Ponty. Noțiunea lui Gibson de „affordances" („ofertele" mediului pentru fiecare dintre noi, n. trad.) – pe care o vom întâlni din nou în capitolul 9 – surprinde ideea că percepem obiectele în termeni de posibilități de comportament pe care le oferă (Gibson, 1979). Husserl a propus că „percepția are orizonturi alcătuite din alte posibilități de percepție, ca percepții pe care le-am putea avea, dacă am dirija în mod activ și altfel cursul percepției" (Husserl, 1960 [1931]). Merleau-Ponty, care a fost puternic influențat de Husserl, a subliniat aspectele corporalizate ale experienței perceptive, a sa *Fenomenologie a percepției,* din 1962, rămânând la fel de importantă (Merleau-Ponty, 1962).

Am propus: Seth (2014b). Acesta este un alt exemplu de „fenomenologie computațională". A se vedea, de asemenea, Cisek (2007) pentru o

Scanned with OKEN Scanner

teorie neurofiziologică conexă în termeni de „competiția ofertei"/ „affordance competition".

sinestezia grafem-culoare: Seth (2014b).

un experiment recent: Suzuki et al. (2019).

Imagine plină de șerpi: un alt exemplu bun este oferit de efectele secundare ale mișcării, cum ar fi iluzia cascadei. Priviți drept la o cascadă (sau la un videoclip al unei cascade) pentru un timp, apoi priviți în altă parte, poate la o stâncă de lângă ea. Suprafața stâncii pare să se miște în sus, în timp stânca rămâne pe loc.

trăirea timpului este halucinație controlată: mulți specialiști în neuroștiințe preocupați de tema timpului nu vor fi de acord cu această idee. De fapt, cele mai multe modele psihologice și neuronale pentru percepția timpului presupun un fel de „pacemaker" implementat neuronal, care servește drept etalon cu care timpul fizic poate fi comparat, dând naștere la percepții ale duratei (van Rijn et al., 2014). Alții cred că percepția timpului depinde de semnalele asemănătoare unui ceas care vin din corp (ritmul cardiac și așa mai departe) – vezi Wittmann (2013) –, dar cercetarea noastră pune la îndoială și această idee (Suárez-Pinilla et al., 2019).

prejudecăți caracteristice: subestimarea duratelor lungi și supraestimarea duratelor scurte este un exemplu de efect de tip „regresie la medie". Acest efect este observat în multe, dacă nu în toate modalitățile perceptive și este o semnătură a inferenței bayesiene, deoarece media poate fi considerată ca o precedentă. În percepția timpului, acest efect este cunoscut sub numele de legea lui Vierordt.

prezentând aceleași prejudecăți: Vezi Roseboom et al. (2019). Un sprijin suplimentar pentru ideea lui Warrick a venit din constatarea că potrivirea dintre modelul de calcul și performanța umană a fost și mai bună atunci când input-ul în rețea a fost limitat la părțile din videoclip la care privea subiectul.

am folosit RMNf: Sherman et al. (2020).

Preferatul meu: Stetson et al. (2007). Evident, la baza macaralei se afla o plasă mare, pentru a prinde voluntarii.

realitatea substituțională: pentru o versiune timpurie a acestui proiect, a se vedea Suzuki et al. (2012).

acum putem testa idei: Phillips et al. (2001). Există implicații sociologice potențial importante pentru această direcție de cercetare. În măsura în care oamenii își experimentează percepțiile ca fiind atât „reale", cât și veridice, va fi dificil de acceptat că alții ar putea avea experiențe perceptuale diferite, chiar și atunci când se confruntă cu aceleași circumstanțe obiective. Acesta este unul dintre motivele pentru care

Scanned with OKEN Scanner

au existat atâtea discuții despre Rochie (vezi capitolul 4). Oamenii care o vedeau într-un fel, pur și simplu nu puteau accepta că sunt posibile și alte moduri de a o vedea, tocmai pentru că și-au trăit percepțiile ca și cum ar dezvălui direct o realitate obiectivă. Acest tip de derivă perceptivă – o generalizare a așa-numitelor echo-chambers din rețelele sociale (mediu sau ecosistem în care participanții găsesc suport pentru convingerile lor personale, prin comunicare repetată a anumitor idei și izolare voită de exterior, n. trad.) – are multe implicații asupra modului în care am putea recunoaște și rezolva sau adapta diferențele dintre indivizi, grupuri și culturi (Seth, 2019c).

După cum a spus Hume: Aceste citate se regăsesc în *Tratatul naturii umane* (1738, 1.3.14.25) și *Investigația privind principiile moralei* (1751, Anexa 1.19) ale lui Hume, așa cum se găsește în Kail (2007), p. 20. Am întâlnit materialul în Dennett (2015).

percepem cu și prin: Filosofii o numesc „transparență" (Metzinger, 2003b).

7. Delir

Până la o treime: Collier (2012).

consecințe grave pe termen lung: Davis et al. (2017).

8. Așteaptă-te pe tine însuți

Paradoxul teletransportului: Acest experiment de gândire a fost atribuit în mod independent filozofului Derek Parfit și autorului Stanislaw Lem.

„snop" de percepții: această viziune asupra sinelui a fost imortalizată în filosofie ca „teoria snopului"/„bundle theory".

carte extraordinară: Metzinger (2003a).

gemenii siamezi pot simți: Acesta este cazul Kristei și Tatianei Hogan, din Columbia Britanică din Canada: www.nytimes.com/2011/05/29/magazine/could-conjoined-twins-share-a-mind.html.

se poate anula: vezi Brugger & Lenggenhager (2014) pentru o trecere în revistă a tulburărilor eului corporal.

în spatele ferestrelor ochilor: Pentru o deconstrucție clasică la persoana întâi a acestei impresii, vezi *On Having No Head* (1961) de Douglas Harding.

piatra de temelie a cercetării: lucrarea originală care descrie iluzia mâinii de cauciuc a fost publicată în 1998 (Botvinick & Cohen, 1998) și

Scanned with OKEN Scanner

a dat naștere unei mici industrii de studii ulterioare. Cercetătorii au verificat dacă iluziile cu trei mâini sau fără mâini sunt posibile, dacă inducerea ideii că am mâini de culori diferite poate modifica prejudecățile rasiale implicite și inclusiv dacă șoarecii sunt susceptibili iluziei „cozii de cauciuc". Vezi Braun et al. (2018) și Riemer et al. (2019) pentru recenzii.

extins la întregul corp: experimentul a funcționat și când voluntarii au văzut spatele unui corp fals, generat de computer – un „avatar" – mai degrabă decât propriul lor corp. Vezi Ehrsson (2007); Lenggenhager et al. (2007).

înscris în istorie și cultură: Monroe (1971).

Am o senzație ciudată: Citat în Tong (2003), p. 105.

Mă văd mințind: Blanke et al. (2004), p. 248.

Factorul comun: Blanke et al. (2015).

dovezi pentru maleabilitatea: Vezi Brugger & Lenggenhager (2014).

iluzia „schimbării corpului cu altcineva": Petkova & Ehrsson (2008).

Scopul BeAnotherLab: www.themachinetobeanother.org.

subliniat de un studiu recent: Lush et al. (2020). În acest articol, folosim mai degrabă termenul de „control fenomenologic", decât „hipnoză", parțial pentru că „hipnoza" poartă un bagaj istoric nefericit. Ideea conform căreia așteptările implicite care decurg din designul experimental pot afecta experiențele și comportamentul participanților este o problemă binecunoscută, deși adesea insuficient apreciată în psihologie, și care poate fi urmărită în timp până la articolul despre așa-numitele „caracteristici ale cererii" (Orne, 1962). Important e că simpla diferențiere dintre atingerea sincronă și cea asincronă nu este suficientă pentru a controla caracteristicile cererii în iluzia mâinii de cauciuc, deoarece oamenii au așteptări mari cu privire la ceea ce vor trăi în aceste condiții diferite (Lush, 2020). Pentru un rezumat al studiilor noastre despre iluzia mâinii din cauciuc, controlul fenomenologic și caracteristicile cererii, a se vedea Seth et al. (2021).

Sugestia hipnotică: Pe lângă inducerea de comportamente și raportări ale experiențelor subiective, sugestia hipnotică poate genera răspunsuri fiziologice și neurofiziologice (sau le poate suprima, în cazul analgeziei hipnotice) (Barber, 1961; M. P. Jensen et al., 2017; Stoelb et al., 2009). Aceste măsuri, aparent mai obiective, pot fi, prin urmare, influențate și de sugestibilitate. Din fericire, toate acestea pot fi privite ca fiind o oportunitate, nu doar ca o problemă. Sugestia hipnotică oferă o metodă solidă de a studia modul în care așteptările de top-down pot genera sau stinge experiența perceptivă. În prezent,

Scanned with OKEN Scanner

desfășurăm o serie de experimente pentru a analiza mai profund care tipuri de experiență perceptivă – a sinelui și a lumii – pot fi modelate sau generate prin efectele sugestiei.

cu adevărat conștient de sine: psihologul estoniano-canadian Endel Tulving a numit acest tip de conștientizare de sine „conștiență autonoetică" (Tulving, 1985).

Jurnalele sale: Înregistrări din jurnal din *Forever Today* de Deborah Wearing (2005).

Clive a avut impresia constantă: Wearing (2005), pp. 202–3.

Nu mai are: www.newyorker.com/magazine/2007/09/24/the-abyss.

percepția socială: existența percepției sociale directe, astfel definită, nu este universal acceptată. Alte abordări propun, în schimb, ideea conform căreia conștientizarea stărilor mentale ale altora este inferată din comportamentul lor, într-o manieră care diferă de percepție. Vezi Gallagher (2008) și C. J. Palmer et al. (2015) pentru mai multe despre aceasta.

Inferența activă în percepția socială: O implicație a inferenței sociale active este aceea că, la fel ca inferența activă vizuală, modelele generative subiacente codifică predicții condiționate despre consecințele acțiunilor. În cazul vederii, așa cum am discutat în capitolul 6, aceste predicții se referă la modul în care s-ar schimba semnalele senzoriale vizuale, cu fiecare acțiune. Acolo, am susținut că aceste predicții condiționale stau la baza proprietății fenomenologice a „obiectualității". Într-o lucrare din 2015, Colin Palmer, Jakob Hohwy și cu mine am sugerat că ceva similar se întâmplă în percepția socială. Ideea noastră este că stările mentale ale altora par „reale" în măsura în care creierul nostru codifică un repertoriu bogat de predicții condiționale despre modul în care se pot schimba, cu fiecare acțiune. De exemplu, o astfel de predicție poate fi despre cum s-ar schimba convingerea sau starea emoțională a cuiva, ca urmare a unei anumite afirmații (cum ar fi „du-te și adu-mi niște vin"). Această idee este echivalentul finisării unei terorii a minții și poate oferi o modalitate utilă de interpretare a deficitelor aparente în percepția socială, așa cum se întâmplă, de exemplu, în autism. Vezi C. J. Palmer et al. (2015).

percepții predictive ierarhizate social: discuțiile neuroștiințifice despre percepția socială evidențiază adesea așa-numiții „neuroni oglindă". Acești neuroni – descoperiți pentru prima dată în creierul maimuțelor de către specialistul italian în neuroștiințe Giacomo Rizzolatti și colegii săi – declanșează impulsuri atunci când un animal acționează, dar și atunci când observă aceeași

Scanned with OKEN Scanner

acțiune efectuată de un alt animal (Gallese et al., 1996). Se spune că neuronii „oglindesc" comportamentul altor animale, deoarece răspund ca și cum observatorul însuși ar acționa. Capacitatea acestor neuroni de a răspunde în acest fel a fost propusă ca fundament pentru toate tipurile de fenomene sociale. Astfel de propuneri pun, totuși, o povară explicativă grea asupra a ceea ce se dovedește, în cele din urmă, a fi doar tipuri specifice de celule cerebrale. Este vorba de aceeași eroare de simplificare pe care o facem atunci când, pe baza scanărilor RMNf, spunem că activitatea dintr-o regiune sau alta explică „dragostea" sau „limbajul". Vezi Caramazza et al. (2014).

niciun om nu este o insulă: „Nimeni nu este o insulă în sine, pe de-a-ntregul; fiecare om este o bucată de continent, o parte a întregului." Donne (1839), pp. 574–5. Psihologul Chris Frith a dus acest punct de vedere și mai departe, susținând că funcția primară a oricărei experiențe conștiente este cea socială (Frith, 2007).

la nivelul identității personale: Antonio Damasio evidențiază acest aspect al sinelui în cartea sa *The Feeling of What Happens* (Damasio, 2000).

Contrar percepției: James (1890), p. 242.

9. A fi un animal-mașină

Nu vedem lucruri: Anaïs Nin, *Seducția Minotaurului* (1961), p. 124. Nin a atribuit remarca unui text talmudic antic.

În Marele Lanț: Conceptul Marelui Lanț al Ființei a apărut în Grecia antică cu Platon, Aristotel și Plotin și s-a dezvotlat pe deplin în vestul Europei în timpul Evului Mediu.

A enerva puternica Biserică Catolică: pentru o relatare istorică animată, vezi *Mașina sufletului* a lui George Makari (2016).

dovediți existența unui Dumnezeu binevoitor: https://en.m.wikipedia.org/wiki/Trademark_argument. Vezi și Hatfield (2002).

Corpurile – atât al omului, cât și al fiarei: Shugg (1968), p. 279. Pentru original, vezi *The Philosophical Works of Descartes,* trad. E. S. Haldane și G. R. T. Ross (New York, 1955), voi. 1, 114–16, 118.

l'homme machine: La Mettrie (1748).

dacă viața și mintea sunt continue: Godfrey-Smith (1996); Maturana & Varela (1980).

simțul internului: Craig (2002).

Semnale senzoriale interoceptive: Critchley & Harrison (2013).

cortexul insular: Vezi Barrett & Simmons (2015) și Craig (2009) pentru mai multe despre rolul cortexului insular în interocepție.

Scanned with OKEN Scanner

Ne pare rău: James (1884), p. 190. Dezbaterea dintre teoreticienii „clasici" ai emoțiilor, care îl urmează pe Darwin atunci când propun emoții înnăscute conservate de la o specie la alta, și „constructiviști", care nu o fac, continuă până în prezent. În fosta tabără îl avem pe biologul Jaak Panksepp și adepții săi. Panksepp a susținut că un set de emoții de bază sunt instanțiate de circuite neuronale specifice (și vechi din punct de vedere evolutiv) (Panksepp, 2004); vezi și Darwin (1872). Cealaltă tabără este reprezentată de cercetătorii în neuroștiință Lisa Feldman Barrett și Joe LeDoux, care propun diferite versiuni ale ideii că emoțiile umane depind de evaluările cognitive – o perspectivă, așa cum vom vedea, similară cu a mea. Pentru mai multe despre istoria teoriilor emoțiilor, a se vedea Barrett (2016); LeDoux (2012).

încă dezbătut: Vezi de exemplu Harrison et al. (2010).

teorii ale evaluării: Schachter & Singer (1962) este referința clasică pentru teoriile evaluării.

un studiu inventiv: Dutton & Aron (1974).

mi-a trecut prin minte: contururile acestei idei au apărut pentru prima dată într-o lucrare din 2011 (Seth et al., 2011b) și au fost rafinate într-un articol din 2013 care, de atunci, a devenit referința standard (Seth, 2013). Ideile de bază au fost extinse începând cu 2015 în teoria „animal-mașină" a conștienței și a sinelui: Seth (2015a); Seth (2019a); Seth & Friston (2016); Seth și Tsakiris (2018).

predicții despre cauzele semnalelor interoceptive: ideile lui Damasio mi-au modelat profund propria gândire, mai ales când vine vorba despre sine; vezi Damasio (1994, 2000, 2010). Lisa Feldman Barrett, ca și mine, subliniază rolul predicțiilor interoceptive în emoție – vezi Barrett & Satpute (2019) și Barrett & Simmons (2015), precum și excelenta sa carte *How Emotions Are Made* (Barrett, 2017).

dificil de testat experimental: Vezi Petzschner et al. (2019). Dovezi suplimentare pentru inferența interoceptivă încep să se acumuleze din studiile pe animale. De exemplu, două experimente recente au sugerat că neuronii din cortexul insular al șoarecilor codifică ceva de genul predicțiilor interoceptive (Gehrlach et al., 2019; Livneh et al., 2020).

sincronism cardio-vizual: Vezi Aspell et al. (2013); Suzuki et al. (2013). Într-un studiu similar, Micah Allen și colegii săi au arătat că excitarea fiziologică neașteptată influențează percepția stimulilor vizuali – sugerând din nou o interacțiune între procesele exteroceptive și interoceptive (Allen et al., 2016). Vezi Park & Blanke (2019) pentru o recenzie, Brener & Ring (2016), Zamariola et al. (2018) pentru discuții despre problemele legate de măsurarea

Scanned with OKEN Scanner

sensibilității interoceptive prin detectarea bătăilor inimii. Influența sugestibilității hipnotice asupra corporalizării a fost discutată în capitolul 8.

studiul științific al controlului: Wiener (1948).

perspective valoroase: istoria ciberneticii arată nu numai modul în care disciplinele academice, pe care acum le considerăm distincte, au fost cândva parte a unei abordări comune, dar dezvăluie și modul în care știința poate, uneori, să se îndepărteze de căile care ar fi putut aduce realizări extraordinare dacă ar fi fost urmate. Pentru mai multe despre aceasta, recomand lucrarea lui Jean-Pierre Dupuy, *Despre Originile Științei Cognitive: Mecanizarea minții* (2009).

„Teorema Bunului Reglator": Conant și Ashby (1970).

trebuie să fie un model: Vezi Seth (2015a) și Seth & Tsakiris (2018) pentru mai multe despre diferența dintre a fi model și a avea un model.

despre a găsi lucruri sau despre a controla lucruri: în articolele despre cercetarea mea, mă refer la diferența dintre ele ca fiind între forme epistemice (exploratorii, de căutare a informației) și instrumentale (orientate spre obiectiv, spre control) de percepție predictivă: Seth (2019a); Seth & Tsakiris (2018); Tschantz et al. (2020b).

variabilă esențială: Ashby (1952). De exemplu, temperatura „miezului" corpului uman trebuie să rămână între 32 °C (89,6 °F) și 40 °C (104 °F), altfel survine moartea.

aceste acțiuni pot fi ambele: Acțiunile externe și cele interne se diferențiază prin tipul de mușchi implicat. Acțiunile externe depind de sistemul muscular scheletic (striat), în timp ce (intero-)acțiunile interne depind de sistemele musculare visceral (neted) și cardiac. Aceste tipuri de mușchi sunt, la rândul lor, controlate de diferite ramuri ale sistemului nervos periferic (parte a sistemului nervos care se află în afara creierului și măduvei spinării). Mușchii scheletici sunt controlați de ramura somatică, iar mușchii viscerali și cardiaci sunt controlați de ramura autonomă.

anularea accelerației optice: Vezi McLeod et al. (2003).

affordances/oferte: Gibson (1979).

Teoria controlului perceptiv: această teorie este adesea rezumată prin sloganul că sistemele de control verifică ceea ce simt, nu ceea ce fac (Powers, 1973). Vezi Marken & Mansell (2013) pentru o formulare mai recentă a teoriei.

Nu există fenomenologie: nu sunt împotriva atribuirii metaforice de forme experiențelor non-vizuale. Durerile pot fi ascuțite sau surde. Unele gusturi sunt, de asemenea, ascuțite, iar unele emoții pot

Scanned with OKEN Scanner

fi descrise și în acest fel – poate gelozia, ca un pumn. Dar aceste experiențe nu au forme în același mod în care cănile, pisicile și măsuțele de cafea experimentate au forme.

allostazie: Sterling (2012). Vezi Tschantz et al. (2021) pentru un model computațional al controlului interoceptiv allostatic.

Teoria animal-mașină: vezi Seth (2013), Seth (2014b), Seth (2015a), Seth (2019), Seth & Friston (2016) și Seth și Tsakiris (2018) pentru mai multe versiuni tehnice ale teoriei beast machine și ale componentelor ei. Teoria are mulți strămoși și influențe, pe care nu am cum să-i menționez pe toți aici. Acestea includ examinarea filozofică a sinelui, făcută de Thomas Metzinger (Metzinger, 2003a) și relatările-reper despre procesarea predictivă făcute de Andy Clark și Jakob Hohwy (Clark, 2016; Hohwy, 2013). Teoria se bazează pe alte propuneri de legături profunde – dar diferite – între viață, corp, minte și conștiență. Aici, am fost puternic influențat de Antonio Damasio (de exemplu, 1994, 2010), Gerald Edelman (de exemplu, 1989), Karl Friston (de exemplu, 2010), Joe LeDoux (de exemplu, 2019) și Evan Thompson (de exemplu, 2014; vezi și Varela et al., 1993). Pentru idei înrudite, vezi Panksepp (2005); Park & Tallon-Baudry (2014); Solms (2021); Conceptul lui Metzinger de „prejudecata existenței" (Metzinger, 2021) și lucrarea Lisei Feldman Barrett (de exemplu, 2017).

animale-mașină conștiente: o asociere strânsă între conștiență și reglarea fiziologică ridică noi întrebări cu privire la rolul trunchiului cerebral – setul de nuclei care se află între cele mai profunde părți ale emisferelor cerebrale și măduva spinării. De obicei, trunchiul cerebral a fost gândit ca un „factor de activare" pentru conștiență, la fel cum un cablu de alimentare este un factor de activare pentru un televizor. Dar trunchiul cerebral joacă un rol extrem de activ în reglarea fiziologică, ceea ce îi face pe unii să sugereze că aici apare conștiența – fără să fie nevoie de cortex (Solms, 2021; vezi și Merker, 2007). Cred că acest lucru este extrem de puțin probabil, având în vedere greutatea dovezilor explicative care leagă cortexul (și talamusul) de stările conștiente. Acestea fiind spuse, trunchiul cerebral poate juca un rol mai important în formarea stărilor conștiente decât sugerează analogia cablului de alimentare – vezi Parvizi & Damasio (2001) pentru o viziune nuanțată.

percepție greșită sistematică: este interesant de luat în considerare dacă orbirea față de autoschimbare este atenuată în timpul bolii sau rănilor, când ar putea fi util pentru creier să perceapă mai precis ceea ce se întâmplă în corp. Există un nou subdomeniu al neuroștiinței

Scanned with OKEN Scanner

cognitive care se ocupă de astfel de întrebări, numit „psihosomatică computațională" (Petzschner et al., 2017).

nu pentru a ne cunoaște pe noi înșine: Grecii realizaseră deja asta. În timp ce Socrate este asociat cu sintagma „cunoaște-te pe tine însuți", stoicii au subliniat importanța calmului și a autocontrolului. Susținătorii teoriei controlului perceptiv ar putea merge mai departe, spunând că ne reglăm starea fiziologică pentru a ne percepe pe noi înșine ca stabili.

o amăgire rară: Cotard (1880).

Semnificativ greșit: o modalitate prin care sinele își poate pierde realitatea este prin incapacitatea modelelor generative subiacente de a codifica un repertoriu bogat de predicții interoceptive condiționale sau contrafactuale cu privire la modul în care acțiunile afectează reglarea fiziologică. Aceasta prin analogie cu modul în care predicțiile vizuale condiționale pot sta la baza fenomenologiei „obiectualității". Vezi Seth & Tsakiris (2018).

ceva special despre materia vie: afirmația conform căreia conștiența depinde de o proprietate specific biologică (pe care un computer din siliciu nu ar putea-o avea niciodată) se numește, uneori, naturalism biologic. Nu folosesc acest termen aici, deoarece a fost folosit în moduri diferite de către diferiți oameni – vezi discuția din Schneider (2019).

10. Un pește în apă

explică toate caracteristicile sistemelor vii: Friston a publicat un număr mare de lucrări despre principiul energiei libere. Două prezentări cheie aparțin lui Friston (2009, 2010).

propria noastră recenzie: Buckley et al. (2017).

se luptă să înțeleagă: www.lesswrong.com/posts/wpZJvgQ4HvJE2bysy/god-help-us-let-s-try-to-understand-friston-on-free-energy. Alte cărți valoroase sunt *„How to read Karl Friston (in the original Greek)"* de Alianna Maren, www.aliannajmaren.com/2017/07/27/how-to-read-karl-friston-in-the-original-greek, și *„Free Energy: How the f*ck does that work, ecologically"* de Andrew Wilson și Sabrina Golonka, psychsciencenotes.blogspot.com/2016/11/free-energy-how-fck-does-that-work.html.

să-și mențină limitele: FEP vorbește despre limite în termeni de „pături Markov" – un concept din statistică și machine learning. Pentru un sistem descris de un set de variabile aleatorii, o pătură

Scanned with OKEN Scanner

Markov este o împărțire statistică a sistemului în „stări interne", „stări externe" și „stări pătură", pătura separând interiorul de exterior. Păturile Markov satisfac cerința conform căreia variabilele din interiorul păturii (stările interne) sunt independente condițional de cele din afara păturii (stările externe) și invers. Aceasta înseamnă că dinamica stărilor interne poate fi pe deplin prezisă din stările interne și stările generale din pătură. Vezi Kirchhoff et al. (2018) pentru mai multe despre păturile Markov și FEP și Bruineberg et al. (2020) pentru o abordare relevantă.

aproximează entropia senzorială: tehnic, energia liberă oferă o limită superioară unei cantități numite surpriză sau auto-informație, care poate fi gândită ca specificând cât de neașteptat (statistic) este un eveniment. Limita superioară înseamnă că energia liberă nu poate fi mai puțin decât surpriza. Surpriza este corelată cu cantitatea teoretică de informație a entropiei prin aceea că, în presupunerile de stare staționară de neechilibru, media pe termen lung a surprizei este entropia. În mod informal, entropia este ca incertitudinea – iar incertitudinea este surpriza medie la care te aștepți.

după niște artificii matematice: Mai multe detalii pentru cei cu înclinație matematică. Energia liberă este definită în termeni de două distribuții de probabilitate: (i) o *densitate de recunoaștere* care codifică cea mai bună estimare curentă despre starea mediului și (ii) o *densitate generativă* care codifică un model probabilistic al modului în care stările de mediu modelează (generează) input-uri senzoriale. Aici „mediu" se referă la cauzele ascunse ale semnalelor senzoriale, oricare ar fi acestea. Cantitățile de energie liberă au două componente: o *energie* care corespunde surprizei și o *entropie relativă* care reflectă cât de „departe" este densitatea de recunoaștere față de adevărata densitate posterioară (probabilitatea stărilor mediului, având în vedere input-urile senzoriale). Distanța dintre aceste densități este măsurată printr-o mărime care, în teoria informației, se numește divergența Kullback-Leibler (KL). Dacă se presupune că recunoașterea și densitățile generative sunt gaussiene (și având în vedere alte presupuneri, de exemplu despre independența intervalelor de timp), atunci energia liberă se potrivește direct pe eroarea de predicție ponderată cu precizie în procesarea predictivă. Deoarece cea mai mică măsură „depărtată" este zero, aceasta înseamnă că energia liberă este întotdeauna mai mare decât surpriză (adică oferă o limită superioară). Acest lucru înseamnă, la rândul său, că scăderea energiei libere trebuie fie să reducă divergența dintre densitatea de recunoaștere și adevăratul posterior (care vine

Scanned with OKEN Scanner

cu inferențe perceptuale mai bune), fie să reducă surpriza (prin prelevarea de noi input-uri senzoriale).

Parafrazându-l pe Friston: Friston (2010). Jakob Hohwy a numit acest proces „auto-evidență" (Hohwy, 2014). Matematic, acest punct de vedere este justificat deoarece minimizarea energiei libere este echivalentă cu maximizarea dovezilor modelului (bayesian). Într-adevăr, așa cum energia liberă oferă o limită superioară pentru surpriză, tot așa oferă o limită inferioară pentru dovezile modelului (așa-numita limită inferioară a dovezilor, sau ELBO, în machine learning); vezi Winn & Bishop (2005).

din camera întunecată: „problema camerei întunecate" a fost una dintre obiecțiile ridicate pentru prima dată împotriva FEP (Friston et al., 2012). A reapărut din nou în timp ce scriam această carte, iar eu și colegii mei am respins-o din nou. Vezi Seth et al. (2020); Sun & Firestone (2020).

FEP va fi judecat: Howwy (2020b), p. 9.

manual de mecanică statistică: Goodstein (1985), p. 1.

mai integrator și mai puternic: un exemplu al acestui mod de a îmbunătăți teoria animal-mașină ne duce înapoi la ideea că sistemele vii mențin o graniță între ele și mediul lor – unde o graniță în FEP este înțeleasă în termenii unei pături Markov. Pentru Friston, simpla existență sau identificarea unei pături Markov implică în mod direct faptul că are loc o inferență activă. Vezi Kirchhoff et al. (2018) și din nou Bruineberg et al. (2020) pentru o critică.

permiteți-ne să facem experimente mai bune: Un exemplu este cercetarea noastră care explorează modul în care agenții de minimizare a energiei libere învață modele de percepție cu prejudecăți adaptative ale mediului lor (Tschantz et al., 2020b).

teorii pentru știința conștiinței: Hohwy & Seth (2020). Au existat și alte încercări de a lega FEP de conștiență, de exemplu în termeni de profunzime temporală a modelelor generative (Friston, 2018); vezi și Solms (2018); Solms (2021); Williford et al. (2018).

încercări în curs de desfășurare: Aceste încercări sunt sub forma unei „colaborări contradictorii", în care susținătorii celor două teorii semnează în avans legat de faptul că rezultatul unui experiment va sprijini sau va submina teoria lor preferată. Această colaborare adversă specifică compară IIT cu inferența activă (nu FEP în sine). Am menționat unul dintre experimentele propuse în capitolul 3: IIT prezice că inactivarea neuronilor deja inactivi va determina o diferență în percepția conștientă, în timp ce inferența activă nu.

Scanned with OKEN Scanner

11. Grade de libertate

Bend the finger (în jargon, a fuma iarbă, n. trad.): McEwan (2000). Îmi ridic pălăria în fața lui Patrick Haggard de la University College London pentru această expresie.

Nu e nici măcar claritate: pentru un ghid priceput al acestui câmp minat filosofic, vezi Bayne (2008).

radical, absolut: Strawson (2008), p. 367.

Poate să tragă în continuare: Această viziune, la care subscriu, exprimă ceea ce filosofii numesc compatibilism. Compatibilismul susține că unele concepții rezonabile despre liberul arbitru sunt compatibile cu ideea de univers determinist. Dimpotrivă, liberul arbitru libertarian (fără nicio legătură cu filosofia politică) este versiunea filosofică a liberului arbitru ciudat. Există, de asemenea, susținători ai așa-numitului determinism dur, după care determinismul este adevărat și, deci, nicio altă concepție rezonabilă despre liberul arbitru nu poate să existe. Rețineți că pot fi agnostic pe tema universului determinist, ceea ce și sunt, și să fiu, totuși, și compatibilist.

nici nu se simt aleatoriu, nici nu sunt: chiar dacă universul este determinist la un nivel fundamental, s-ar putea ca fluctuațiile aparent aleatorii la nivelul neuronilor și al sinapselor să joace un rol semnificativ în funcționarea creierului. Acest lucru este posibil, poate chiar probabil, dar, din nou, nu contează.

fenomen binecunoscut: potențialul pre-motor a fost înregistrat pentru prima dată în anii 1960 de către fiziologii germani Hans Kornhuber și Lüder Deecke, care l-au numit Bereitschaftpotential (Kornhuber & Deecke, 1965).

a simțit „nevoia": Pentru filosofi, intențiile și nevoile sunt lucruri diferite. S-ar putea să simt nevoia de a lovi pe cineva care mă enervează, dar voi inhiba acest impuls pentru că nu intenționez să rănesc pe nimeni sau să fiu arestat. Intențiile sunt supuse unor judecăți și norme într-un mod în care nevoile nu sunt. În cazuri simple, cum ar fi îndoirea voluntară a unui deget, ele reprezintă aproape același lucru, iar Libet însuși a folosit în mare parte termenii în mod interschimbabil.

a început deja: Libet (1985). Deloc surprinzător, atât debutul potențial de pregătire, cât și momentul intenției conștiente au loc înainte de mișcarea în sine.

free won't (inversul lui free will/liber arbitru, n. trad.): Libet et al. (1983).

nu o vedem: Filosoful Al Mele enunțase un concept similar cu câțiva ani mai devreme (Mele, 2009).

Scanned with OKEN Scanner

mic semn de orice: Schurger et al. (2012).

Dacă sunt forțat: Vezi Caspar et al. (2016) pentru un experiment interesant pe această temă.

Omul poate face ce dorește: Acest citat provine dintr-un eseu pe care Schopenhauer l-a prezentat Societății Regale Norvegiene de Științe în 1839. Pentru o traducere, vezi Zucker (2013), p. 531. În varianta lui Schopenhauer se pot vedea rădăcinile oricărei dependențe.

grade de libertate: Există o legătură interesantă cu cibernetica aici. Ross Ashby, pe care l-am întâlnit în capitolul 9 cu ale sale Teoremă a Bunului Reglator și concept de variabilă esențială, este cunoscut și pentru mai timpuria Lege a Varietății Necesare (Ashby, 1956). Această lege, sau principiu, afirmă că un sistem de control de succes trebuie să fie capabil să intre în cel puțin atâtea stări câte ia mediul care îl perturbă. Așa cum a spus Ashby, „numai varietatea poate împinge în jos varietatea". Vezi Seth (2015a).

Comportamentul voluntar depinde: Dennett (1984).

trei procese: Brass & Haggard (2008); Haggard (2008); Haggard (2019).

stimularea acestor regiuni: Fried et al. (1991). O stimulare mai intensă a acestei zone poate genera atât o nevoie, cât și acțiunea corespunzătoare.

localizabil la părți mai frontale: Brass & Haggard (2007).

Iluzia Voinței Conștiente: Wegner (2002).

o problemă care nu există: așa cum a spus Sam Harris într-un podcast recent, „nu numai că problema liberului arbitru nu are sens obiectiv, dar nu are nici subiectiv." A se vedea https://samharris.org/podcasts/241-final-thoughts-on-free-will/.

fiind „în moment": Csikszentmihalyi (1990); Harris (2012).

Leziuni ale creierului sau neșansă: Della Sala et al.(1991); Formisano et al. (2011).

Tumora cerebrală localizată ciudat: cazul pedofiliei induse de tumori este descris în Burns & Swerdlow (2003). Cel al lui Charles Whitman a fost relatat de multe ori, inclusiv într-un articol din 2011 din *The Atlantic* de David Eagleman: www.theatlantic.com/magazine/archive/2011/07/the-brain-on-trial/308520.

Einstein a declarat: „Nu-mi asum niciun merit pentru nimic. Totul este determinat, atât începutul, cât și sfârșitul, de forțe asupra cărora nu avem niciun control. Este determinat atât pentru insectă, cât și pentru stea. Ființe umane, legume sau praf cosmic, toți dansăm pe o melodie misterioasă." În acest pasaj ni se relevă dedicarea totală a lui Einstein față de ideea de determinism. (Tot el a afirmat, într-un citat celebru, că „Dumnezeu nu joacă zaruri cu universul" – respingând

Scanned with OKEN Scanner

aleatoriul inerent mecanicii cuantice.) Cu toate acestea, după cum am văzut, nu este nevoie să îmbrățișăm determinismul pentru a respinge liberul arbitru ciudat. Pasajul provine dintr-un interviu realizat de George Sylvester Viereck, publicat în *Saturday Evening Post* la 26 octombrie 1929 (p. 117).

precum filozoful Bruce Waller: Vezi Waller (2011) pentru opinia conform căreia responsabilitatea morală este incoerentă, și Dennett (1984, 2003) pentru alternativă. Există o dezbatere recentă revelatoare cu privire la această problemă în Dennett & Caruso (2021).

printre animale: Neurobiologul Björn Brembs a susținut că urme ale ceea ce numim liber arbitru pot fi găsite chiar și la creaturi foarte simple, atunci când se comportă în moduri variabile și aparent spontane care par să vină „din interior". Deoarece aceste comportamente – cum ar fi „răspunsul de evadare" imprevizibil al unui gândac – sunt benefice pentru evitarea prădătorilor, poate reflectă originile evolutive ale modului în care noi, oamenii, am ajuns să ne controlăm multele grade de libertate. Vezi Brembs (2011).

12. Dincolo de om

istoricul urmăririi penale a animalelor: Evans (1906).

Animale-robot ale dualismului cartezian: opiniile lui Descartes despre animale au fost publicate în secolul al XVII-lea coexistând, astfel, parțial cu aceste credințe și practici medievale. Viziunea carteziană a devenit dominantă mai târziu, pe măsură ce Iluminismul a luat în stăpânire toată Europa.

absența așa-numitelor abilități cognitive „de nivel înalt": Avocații teoriilor „gândirii de ordin superior" ale conștienței (pe care le-am descris pe scurt în capitolul 1) ar putea să nu fie de acord cu asta. Vezi Brown et al. (2019).

animalele ca animal-mașină: La extrema opusă față de Descartes, Darwin a adoptat o perspectivă puternic antropomorfă, în special în ceea ce privește exprimarea emoțiilor la animale (Darwin, 1872). Presupunerea sa privind un set conservat de emoții pentru multe specii a autorizat utilizarea experimentelor pe animale pentru a studia emoțiile umane, direcție de lucru preluată mai târziu de Jaak Panksepp și alții, și s-a transmis în cultura contemporană sub formă de emoții „hard-wired (instinctive, n. trad.)" sau „de bază", asociate cu expresii faciale specifice, ca în lucrarea lui Paul Ekman (1992). După cum am explicat în capitolul 9, această viziune a „emoției de bază" a fost contestată de „constructiviștii" contemporani, precum

Scanned with OKEN Scanner

Lisa Feldman Barrett și Joe LeDoux care, ca și mine, subliniază rolul interpretărilor top-down în formarea conținutului conștient. Vezi Barrett (2016); LeDoux (2012).

Chiar și șobolanii, sugerează un studiu: Vezi Steiner & Redish (2014), p. 1001.

oriunde există viață: biologul Ernst Haeckel a formulat termenul de „biopsihism" în 1892 pentru a descrie concepția conform căreia toate și numai vietățile simt – spre deosebire de viziunea panpsihistă conform căreia conștiența este o proprietate a tuturor formelor materiei. Vezi Thompson (2007).

lăsând deoparte dimensiunea brută a creierului: o măsură mai sofisticată decât dimensiunea brută a creierului este coeficientul de encefalizare, o măsură a dimensiunii relative a creierului, care ia în considerare dimensiunea corpului. Există multe discuții pe tema considerării coeficientului de encefalizare ca predictor fiabil al abilității cognitive la mai multe specii (Herculano-Houzel, 2016; Reep et al., 2007). Cu toate acestea, sunt prea puține motivele pentru a considera fie dimensiunea creierului, fie coeficientul de encefalizare ca marker al prezenței conștienței la mai multe specii.

șaptesprezece proprietăți distincte: Seth et al. (2005).

efecte similare la speciile de mamifere: Kelz & Mashour (2019).

foci... și delfini: Lyamin et al. (2018); Walker (2017).

univers interior distinctiv: Uexküll (1957).

dezvoltat de psihologul Gordon Gallup Jr: Gallup (1970).

niște maimuțe antropoide: cum se precizează într-o articol cu unul dintre titlurile mele mereu preferate: „O altă gorilă *(Gorilla gorilla gorilla)* se recunoaște în oglindă" (Posada & Colell, 2007).

nicio dovadă convingătoare: pentru o sinteză, a se vedea Gallup & Anderson (2020). Dezbaterea despre peștii-doctor este relatată în Kohda et al. (2019) și de Waal (2019).

Dognition (echivalentul cunoașterii, pentru câini, n. trad.): Gallup & Anderson (2018).

maimuțele au fost chiar dresate să: Cowey & Stoerig (1995).

un echivalent primat: Boly et al. (2013).

transplantat acolo în 1938: Kessler & Rawlins (2016).

Într-un videoclip: www.ted.com/talks/frans_de_waal_do_animals_have_morals.

Nu a trecut testul oglinzii: Un studiu a constatat că maimuțele rhesus pot trece testul oglinzii după câteva săptămâni de instruire (L. Chang et al., 2017). Dar trecerea testului după un antrenament extins este foarte diferită de utilizarea spontană a oglinzilor pentru autorecunoaștere.

Scanned with OKEN Scanner

A Catalog of Body Patterning in Cephalopoda: Borrelli et al. (2006). Cartea asta minunată, dar extrem de grea, a făcut să depășesc complet maximul de greutate permis de Ryanair.

Mintea unui extraterestru: În 2015, am fost invitat să scriu un capitol de carte despre „conștiența extraterestră". După ce m-am agitat o vreme, am decis să scriu despre caracatițe, în loc să speculez despre extratereștri potențiali existenți (Seth, 2016a). Paralela dintre caracatițe și extratereștri a fost explorată și în filmul artistic din 2016 al lui Denis Villeneuve, *Arrival.*

Dacă vrem să înțelegem alte minți: Godfrey-Smith (2017), p. 10.

creierii de caracatiță sunt lipsiți de mielină: Hochner (2012); Shigeno et al. (2018).

Conștiența caracatiței: Carls-Diamante (2017).

Unii cercetători au sugerat: Liscovitch-Brauer et al. (2017).

abilitățile lor cognitive: Fiorito & Scotto (1992). Vezi și D. B. Edelman & Seth (2009); Mather (2019); și, pentru un text clasic despre comportamentul cefalopodelor, Hanlon & Messenger (1996).

caracatiță surprinsă în câmp deschis: www.bbc.co.uk/programmes/p05nzfn1. Există imagini la fel de impresionante în documentarul *My Octopus Teacher* din 2020 de Pippa Ehrlich și James Reed, ce prezintă un regizor – Craig Foster – care dezvoltă o relație surprinzător de strânsă cu o caracatiță.

se amestecă în fundal: biologul marin Roger Hanlon a înregistrat video multe exemple de camuflaj la caracatiță. Iată una dintre cele mai bune: www.youtube.com/watch?v=JSq8nghQZqA.

creierul central poate nici să nu știe: Mather (2019); Messenger (2001).

caracatițele pot gusta cu ventuzele: van Giesen et al. (2020).

experimente macabre: Nesher et al. (2014).

animale pentru care „lumina e aprinsă": Vezi D. B. Edelman & Seth (2009) și D. B. Edelman et al. (2005) pentru mai multe despre conștiența non-mamiferă.

Multe specii de păsări: Clayton et al. (2007); Jao Keehn et al. (2019); Pepperberg & Gordon (2005); Pepperberg & Shive (2001).

probabil să aibă experiențe conștiente: interesant, creierul păsărilor are ceva asemănător cortexului mamiferelor (numit „pallium"), dar le lipsește orice echivalent al corpului calos, care conectează, la mamifere, cele două emisfere corticale. Prin urmare, păsările ar putea reprezenta un fel de „creier divizat natural", ridicând întrebări despre unitatea conștienței aviare (Xiao & Gunturkun, 2009). *The Thing with Feathers* (2014) de Noah Strycker este o introducere minunată în cunoașterea și comportamentul aviar.

Scanned with OKEN Scanner

Pe măsură ce ne deplasăm mai departe: Pentru abordări excelente ale evoluției conștienței, vezi Feinberg & Mallatt (2017); Ginsburg & Jablonka (2019); LeDoux (2019).

Deciziile privind bunăstarea animalelor: Vezi Birch (2017) pentru o prezentare generală utilă, motivată de argumentul că ar trebui să oferim animalelor „prezumția de nevinovăție" atunci când dovezile sunt neconcludente (un argument cunoscut mai formal ca „principiul precauției"). De remarcat că Uniunea Europeană a decis în 2010 să includă cefalopodele în legislația privind bunăstarea animalelor (directiva 2010/63/UE).

creierul de insectă chiar are: Entler et al. (2016).

musculița de oțet: Khuong et al. (2019).

eficient pentru toate animalele: Kelz & Mashour (2019).

posibile minți conștiente: Jonathan Birch, Alexandra Schnell și Nicola Clayton sugerează în mod util termenul „profil de conștiență" pentru a caracteriza variația experienței conștiente la diverse specii. Ei propun cinci dimensiuni ale variației: bogăția perceptivă, bogăția evaluativă, unitatea, temporalitatea și individualitatea. Vezi Birch et al. (2020).

13. Minți-mașină

un golem: în cartea sa din 1964, *God and Golem, Inc.*, pionierul polimat Norbert Wiener a tratat golemii ca fiind esențiali pentru speculațiile sale despre riscurile viitoarei AI.

grămadă mare de agrafe: în parabola maximizatorului de agrafe, un AI este conceput pentru a face cât mai multe agrafe posibil. Deoarece această inteligență artificială nu are valori umane, dar este, însă, foarte inteligentă, ajunge să distrugă lumea în încercarea sa reușită în acest sens. Vezi Bostrom (2014).

așa-numita ipoteză a „Singularității": Vezi Shanahan (2015) pentru o interpretare revigorant de sobră a ipotezei Singularității.

inteligența poate exista fără conștiență: este tentant să afirmăm că inteligența și conștiența sunt dublu disociabile – că fiecare poate exista fără cealaltă. Dar nu ar fi tocmai corect. Deși cred că inteligența poate exista fără conștiență, s-ar putea ca, totuși, conștiența să necesite un anumit nivel de inteligență.

nu o singură scală: ideea de conștiență (și inteligență) multidimensională amintește de conceptul lui Jonathan Birch și al colegilor săi despre un „profil de conștiență" (Birch et al., 2020), precum și abordarea multidimensională a nivelurilor de conștiență la oameni propusă de

Scanned with OKEN Scanner

Tim Bayne, Jakob Hohwy și Adrian Owen (Bayne et al., 2016).

Autorii se exprimă echivoc: Dehaene et al. (2017). Disponibilitatea globală corespunde teoriei populare a spațiului de lucru global a conștienței, în timp ce „auto-monitorizarea" surprinde aspecte ale teoriilor gândirii de ordin superior. Am întâlnit pe fugă ambele teorii în capitolul 1. Autorii articolului din Science admit în mod explicit că ar putea omite componenta „experiențială" a conștienței. Pentru mine, asta înseamnă a omite prea mult.

mașinile ar putea părea conștiente: această posibilitate apare deoarece IIT acceptă o parte a funcționalismului (independența substratului), dar nu și cealaltă (suficiența mapărilor de input-output). Unele mecanisme, în special rețelele neuronale artificiale feedforward de dimensiune adecvată, pot implementa mapări input-output arbitrar complexe. Aceste mecanisme, implementate în mod corect, ar putea da aspectul exterior al inteligenței și/sau al conștienței. Dar rețelele pur feedforward nu generează deloc informații integrate – este întotdeauna nevoie de o anumită recurență sau „reluare în buclă". Prin urmare, IIT acreditează conceptul de „zombi comportamental", care, așa cum am explicat în capitolul 1, este un artefact ce pare conștient din exterior, dar nu are conștiență. Vezi Tononi & Koch (2015).

regenerând continuu condițiile: Aceste idei se referă îndeaproape la conceptul de autopoieză/autopoiesis (din elină: auto = „sine" și poiesis = „creație"), dezvoltat de biologul chilian Humberto Maturana. Un sistem autopoietic este capabil să se mențină și să se reproducă singur, ceea ce include producerea componentelor fizice necesare existenței sale continui, ca sistem. Deși autopoieza este, în primul rând, o teorie a celulei, există legături interesante între autopoieza celulară și principiul energiei libere (vezi capitolul 10). Ambele sugerează o continuitate solidă între „viață" și „minte", ceea ce sugerează, la rândul ei, că mintea (și, de asemenea, conștiența) se ocupă de mai mult decât ceea ce „face", pur și simplu, un sistem (Kirchhoff, 2018; Maturana și Varela, 1980). Am avut norocul să-l cunosc pe Maturana – care a murit în mai 2021 la vârsta de nouăzeci și doi de ani – în ianuarie 2019, în orașul său natal, Santiago, unde am petrecut timp savurând cafea și discutând aceste idei la o terasă umbrită din Barrio Providencia.

Testul Turing: În varianta originală, „jocul imitației" al lui Alan Turing implică doi oameni de același sex și o mașină. Și mașina și unul din oameni – colaboratorul – pretind că sunt un om de sex opus. Celălalt om trebuie să decidă care este mașina și care este colaboratorul (Turing, 1950).

Scanned with OKEN Scanner

Testul Garland: Acest termen a fost formulat de Murray Shanahan, a cărui carte, *Embodiment and the Inner Life* (2010), a fost una dintre inspirațiile pentru *Ex Machina.*

proclamații zgomotoase: www.reading.ac.uk/news-archive/press-releases/pr583836.html.

Când chatbot-ul a câștigat: „Eugene Goostman este un băiat adevărat – testul Turing spune asta", notează *Guardian Pass,* 9 iunie 2014. A se vedea https://www.theguardian.com/technology/shortcuts/2014/jun/09/eugene-goostman-turing-test-program-calculator.

oamenii au eșuat: descrierea testului Turing ca unul al „credulității umane" provine dintr-un articol din 2015 din *New York Times* al lui John Markoff, „Software-ul este suficient de inteligent pentru SAT, dar încă departe de a fi inteligent", *New York Times,* 21 septembrie 2015. A se vedea www.nytimes.com/2015/09/21/technology/personaltech/software-is-smart-enough-for-sat-but-still-far-from-intelligent.html.

vastă rețea neuronală artificială: GPT înseamnă „Generative Pre-trained Transformer" – un tip de rețea neuronală specializată pentru predicția și generarea limbajului. Aceste rețele sunt antrenate cu ajutorul unei abordări deep learning nesupravegheată, în esență pentru a „prezice următorul cuvânt"¨, dat fiind un cuvând sau un fragment de text anterior. GPT-3 are uimitorul număr de 175 de miliarde de parametri și a fost antrenat pe aproximativ 45 de terabytes de date text. Vezi https://openai.com/blog/openai-api/ și pentru detalii tehnice: https://arxiv.org/abs/2005.14165.

nu înțelege: Desigur, asta depinde de ce anume înseamnă „înțelegere". Unii ar putea afirma că „înțelegerea" umană nu este diferită, ca natură, de acea „înțelegere" expusă de GPT-3. Gary Marcus, specialist în științe cognitive, argumentează împotriva acestei idei și sunt de acord cu el. A se vedea www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/.

un eseu de cinci sute de cuvinte: „Un robot a scris acest articol în întregime. Ți-e deja frică, omule?", *Guardian Opinion,* 8 septembrie 2020. Consultați www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3. Nu este clar cât de reprezentativ este acest exemplu.

cel mai frecvent sentiment... a fost frica: Becker-Asano et al. (2010).

teorii despre motivul pentru care există valea stranie: Mori et al. (2012).

Tehnologii „deepfake": „deepfake" înseamnă a genera un videoclip realist, dar fals, de obicei cu o față umană, folosind machine learning

Scanned with OKEN Scanner

pentru a combina o sursă și un videoclip țintă. Într-un exemplu răspândit pe scară largă din 2017, metoda deepfake a fost folosită pentru a crea videoclipuri convingătoare cu Barack Obama spunând lucruri pe care nu le-a spus niciodată (https://www.youtube.com/watch?v=cQ54GDm1eL0). O serie de videoclipuri TikTok care l-au deepfake pe Tom Cruise, lansate în 2021, ridică ștacheta substanțial (https://www.theverge.com/22303756/tiktok-tom-cruise-impersonator-deepfake).

vast experiment global necontrolat: cercetătorul AI Stuart Russell descrie elocvent amenințările care vin din partea AI actuală și viitoare, precum și modalități de a reproiecta sistemele AI pentru a evita aceste amenințări, în cartea sa *Human Compatible* (2019). Nina Schick face o treabă la fel de bună pentru amenințarea reprezentată de deepfake (Schick, 2020).

instrumente inteligente, nu colegi: „Philosopher Daniel Dennett on AI, robots and religion", *Financial Times,* 3 martie 2017. A se vedea https://www.ft.com/content/96187a7a-fce5-11e6-96f8-3700c5664d30.

moratoriu imediat de treizeci de ani: Vezi Metzinger (2021).

creând noi forme de viață: Emmanuelle Charpentier și Jennifer Doudna au câștigat Premiul Nobel pentru Chimie în 2020 pentru contribuția lor la dezvoltarea tehnicii CRISPR. *E. coli* sintetică a fost creată în laboratorul lui Jason Chin; vezi Fredens et al. (2019).

unde coordonate de activitate electrică: Trujillo et al. (2019).

foarte puțin probabil să fie conștient: am investigat posibilitatea conștienței organoide într-un articol recent cu Tim Bayne și Marcello Massimini (Bayne et al., 2020).

Urgență etică: Aceste probleme sunt luate în serios. În vara anului 2020, am fost invitat – împreună cu alți câțiva specialiști în neuroștiință – să vorbesc unui comitet mixt al Academiei Naționale din SUA, reunit pentru a ajuta la stabilirea cadrelor de reglementare și legale referitoare la cercetarea ce implică atât organoizi, cât și himere (animale modificate genetic pentru a exprima caracteristici umane specifice). Consultați www.nationalacademies.org/our-work/ethical-legal-and-regulatory-issues-associated-with-neural-chimeras-and-organoids.

Vrem să facem ferme: Carl Zimmer, „Organoids are not brains. How are they making brain waves?", *New York Times,* 29 august 2019. A se vedea www.nytimes.com/2019/08/29/science/organoids-brain-alysson-muotri.html.

un trop preferat al futuriștilor: pentru o discuție echilibrată despre

Scanned with OKEN Scanner

perspectivele și capcanele mind uploading/încărcării minții (pe procesoare, n. trad.) vezi Schneider (2019).

agenți virtuali care simt: argumentul simulării se desfășoară astfel (Bostrom, 2003). O civilizație dintr-un viitor suficient de îndepărtat, care a evitat propria extincție, ar avea probabil acces la resurse computaționale vaste. Unii membri ai acestei civilizații pot alege să ruleze pe computer simulări detaliate ale strămoșilor lor. Având în vedere că s-ar putea rula o mulțime de astfel de simulări, ar fi, deci, logic ca orice individ care trăiește acum să concluzioneze că are mai multe șanse să se afle printre mințile simulate decât printre oamenii biologici originali. După cum spune Bostrom: „dacă nu credem că trăim acum într-o simulare pe computer, nu avem dreptul să credem că vom avea descendenți care vor rula o mulțime de astfel de simulări ale strămoșilor lor" (Bostrom, 2003, p. 243). Una dintre multele probleme pe care le am cu acest argument este că presupune despre funcționalism că este adevărat: că, atunci când vine vorba de conștiență, simularea este echivalentă cu instanțierea. După cum am menționat anterior, nu cred că funcționalismul este o presupunere sigură.

Epilog

o insulă de cortex vie, dar izolată: Deoarece este încă conectată la alimentarea cu sânge și „vie", ar putea această emisferă deconectată să-și susțină propria conștiență izolată? Potențiale „insule ale conștientizării" ca aceasta ar putea apărea și în alte neuro-tehnologii emergente, cum ar fi creierii de porc reanimați ex cranio și organoizii cerebrali pe care i-am descris în capitolul anterior. Discutăm toate aceste cazuri în Bayne et al. (2020).

Este larg acceptat: Chalmers (1995b), p. 201.

aspecte ale percepției: Vezi și Hoffman (2019).

Scanned with OKEN Scanner

Referințe

Albright, T. D. (2012). 'On the perception of probable things: neural substrates of associative memory, imagery, and perception'. *Neuron*, 74(2), 227–45.

Allen, M., Frank, D., Schwarzkopf, D. S., et al. (2016). 'Unexpected arousal modulates the influence of sensory noise on confidence'. *Elife*, 5, e18103.

Anscombe, G. E. M. (1959). *An Introduction to Wittgenstein's Tractatus*. London: St. Augustine's Press.

Aru, J., Bachmann, T., Singer, W., et al. (2012). 'Distilling the neural correlates of consciousness'. *Neuroscience and Biobehavioral Reviews*, 36(2), 737–46.

Ashby, W. R. (1952). *Design for a Brain*. London: Chapman and Hall.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman and Hall.

Aspell, J. E., Heydrich, L., Marillier, G., et al. (2013). 'Turning the body and self inside out: Visualized heartbeats alter bodily self-consciousness and tactile perception'. *Psychological Science*, 24(12), 2445–53.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. New York, NY: Cambridge University Press.

Barber, T. X. (1961). 'Physiological effects of "hypnosis"'. *Psychological Bulletin*, 58, 390–419.

Barnes, J. (2008). *Nothing to Be Frightened of*. New York, NY: Knopf.

Barnett, L., Muthukumaraswamy, S. D., Carhart-Harris, R. L., et al. (2020). 'Decreased directed functional connectivity in the psychedelic state'. *Neuroimage*, 209, 116462.

Barrett, A. B., & Seth, A. K. (2011). 'Practical measures of integrated information for time-series data'. *PLoS Computational Biology*, 7(1), e1001052.

Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Boston, MA: Houghton Mifflin Harcourt.

Barrett, L. F., & Satpute, A. B. (2019). 'Historical pitfalls and new directions in the neuroscience of emotion'. *Neuroscience Letters*, 693, 9–18.

Scanned with OKEN Scanner

Boly, M., Seth, A. K., Wilke, M., et al. (2013). 'Consciousness in humans and non-human animals: recent advances and future directions'. *Frontiers in Psychology*, 4, 625.
Borrelli, L., Gherardi, F., & Fiorito, G. (2006). *A Catalogue of Body Patterning in Cephalopoda*. Florence: Firenze University Press.
Bostrom, N. (2003). 'Are you living in a computer simulation?' *Trends in Cognitive Sciences*, 53(11), 243–55.
Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
Botvinick, M., & Cohen, J. (1998). 'Rubber hands "feel" touch that eyes see'. *Nature*, 391(6669), 756.
Brainard, D. H., & Hurlbert, A. C. (2015). 'Colour vision: understanding #TheDress'. *Current Biology*, 25(13), R551–4.
Brass, M., & Haggard, P. (2007). 'To do or not to do: the neural signature of self-control'. *Journal of Neuroscience*, 27(34), 9141–5.
Brass, M., & Haggard, P. (2008). 'The what, when, whether model of intentional action'. *Neuroscientist*, 14(4), 319–25.
Braun, N., Debener, S., Spychala, N., et al. (2018). 'The senses of agency and ownership: a review'. *Frontiers in Psychology*, 9, 535.
Brembs, B. (2011). 'Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates'. *Proceedings of the Royal Society B: Biological Sciences*, 278(1707), 930–39.
Brembs, B. (2020). 'The brain as a dynamically active organ'. *Biochemical and Biophysical Research Communications*. doi:10.1016/j.bbrc.2020.12.011.
Brener, J., & Ring, C. (2016). 'Towards a psychophysics of interoceptive processes: the measurement of heartbeat detection'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160015.
Brown, H., Adams, R. A., Parees, I., et al. (2013). 'Active inference, sensory attenuation and illusions'. *Cognitive Processing*, 14(4), 411–27.
Brown, R., Lau, H., & LeDoux, J. E. (2019). 'Understanding the higher-order approach to consciousness'. *Trends in Cognitive Sciences*, 23(9), 754–68.
Brugger, P., & Lenggenhager, B. (2014). 'The bodily self and its disorders: neurological, psychological and social aspects'. *Current Opinion in Neurology*, 27(6), 644–52.

Scanned with OKEN Scanner

Main: MIND Group, 35(R).
Seth, A. K. (2016a). 'Aliens on earth: What octopus minds can tell us about alien consciousness'. In J. Al-Khalili (ed.), *Aliens*, London: Profile Books, 47–58.
Seth, A. K. (2016b). 'The real problem'. *Aeon*. aeon.co/essays/the-hard-problem-of-consciousness-is-a-distraction-from-the-real-one.
Seth, A. K. (2017). 'The fall and rise of consciousness science'. In A. Haag (Ed.), *The Return of Consciousness* (pp. 13–41). Riga: Ax:Son Johnson Foundation.
Seth, A. K. (2018). 'Consciousness: The last 50 years (and the next)'. *Brain and Neuroscience Advances*, 2, 2398212818816019.
Seth, A. K. (2019a). 'Being a beast machine: The origins of selfhood in control-oriented interoceptive inference'. In M. Colombo, L. Irvine, & M. Stapleton (eds), *Andy Clark and his Critics*, Oxford: Wiley-Blackwell, 238–54.
Seth, A. K. (2019b). 'From unconscious inference to the Beholder's Share: Predictive perception and human experience'. *European Review*, 273(3), 378–410.
Seth, A. K. (2019c). 'Our inner universes'. *Scientific American*, 321(3), 40–47.
Seth, A. K., Baars, B. J., & Edelman, D. B. (2005). 'Criteria for consciousness in humans and other mammals'. *Consciousness and Cognition*, 14(1), 119–39.
Seth, A. K., Barrett, A. B., & Barnett, L. (2011a). 'Causal density and integrated information as measures of conscious level'. *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, 369(1952), 3748–67.
Seth, A. K., Dienes, Z., Cleeremans, A., et al. (2008). 'Measuring consciousness: relating behavioural and neurophysiological approaches'. *Trends in Cognitive Sciences*, 12(8), 314–21.
Seth, A. K., & Friston, K. J. (2016). 'Active interoceptive inference and the emotional brain'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160007.
Seth, A. K., Izhikevich, E., Reeke, G. N., et al. (2006). 'Theories and measures of consciousness: An extended framework'. *Proceedings of the National Academy of Sciences of the USA*, 103(28), 10799–804.
Seth, A. K., Millidge, B., Buckley, C. L., et al. (2020). 'Curious inferences: Reply to Sun and Firestone on the dark room problem'. *Trends in Cognitive Sciences*, 24(9), 681–3.

Scanned with OKEN Scanner

Anil Seth este specialist în neuroștiință, autor și prezentator, a cărui muncă de pionierat în cercetarea bazelor cerebrale ale conștienței este cunoscută de peste 20 de ani. Este autor al cărții *A fi tu: o nouă știință a conștienței* (septembrie 2021) și al best-seller-ului *30 Seconds Brain*, precum și al altor cărți. A fost invitat la TED, în 2017, prelegerea sa despre conștiență ajungând să fie vizualizată de peste 13 milioane de ori, a apărut în mai multe filme (*The Most Unknown*, *The Search*) și a semnat mai multe articole pentru *Aeon*, *The Guardian*, *Granta*, *New Scientist*, *Scientific American*. În 2017, a fost președintele *British Science Association* (secțiunea Psihologie). În 2019, a primit premiul *KidSpirit Perspectives*. A publicat peste 200 de articole academice și, din 2019, se află pe lista *Web of Science* a celor mai citați cercetători, printre cei mai influenți oameni de știință ai deceniului trecut.

„O sinteză amplă care reunește componente disparate din filosofie, știință, literatură, experiență personală și speculație – aceasta din urmă fiind cea mai incitantă pentru mine, în ciuda faptului că unele propuneri nu au fost încă dovedite. Seth își propune să explice nu doar ce și cum suntem, ci – probabil provocator pentru unii – de ce suntem așa cum suntem. De ce avem sentimentul că suntem în permanență aceeași persoană? (Când, evident, eu, cel puțin, nu sunt.) De ce avem acest sentiment de a fi conștienți de sine? Pentru ce este? Extrem de inspirator."

David Byrne

„Cu adevărat convingător... Modul de abordare al conștienței este eclectic și livrat cu un anumit tip de generozitate... O relatare de forță a simțirii și sinelui întrupate. Relatare devenită irezistibilă prin argumentele echilibrate și incluzive ale autorului."

Profesor Karl Friston, University College London (cel mai citat specialist în neuroștiință)

„O carte fascinantă. O bucurie să o citești. Anil Seth explorează întrebări fundamentale despre conștiență și sine din perspectiva unui neurocercetător instruit în filosofie. Recomand cu căldură."

Nigel Warburton, autorul cărții *A Little History of Philosophy*

„O relatare minunat de accesibilă și cuprinzătoare despre modul în care mintea noastră sintetizează lumea și despre cum anume ne face asta ceea ce suntem."

Sean Carroll, autorul cărții *Something Deeply Hidden*

ISBN: 978-973-0-38402-4

Scanned with OKEN Scanner